“Un-Human Minds: Why Being Human is Nothing Special”

By Jack Lowry

The Undergraduate Research Writing Conference

• 2020 •

Rutgers, The State University of New Jersey

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

Jack Lowry

Honors Exposition and Argument

Professor Reagan Lothes

11/30/19

Un-Human Minds: Why Being Human is Nothing Special

If I were to ask someone to define what consciousness meant to them, they would probably respond with a vague answer that looks something along the lines of "being aware, having the ability to react to outside stimuli and think about your decisions". However, most everyone would agree in saying that humans are conscious, and that animals like dogs don't have consciousness, or at least not to the extent that humans have. And yet, dogs do indeed have all the qualities described in their definition; they are aware of their surroundings, can respond to outside stimuli, and seem to take time to think about their decisions. However, many would still argue that we are appreciably different from dogs in terms of the ways we think. Furthermore, if the dog were replaced with an advanced robot of some kind, one's defense of their own consciousness would grow ardently. If our differences do not lie in consciousness alone, then we must consider instead the container for all thought, or mental faculties: our minds. So, what makes our mind so special, and why are we so hasty to defend it when compared to other, so-called lesser minds?

The topic of consciousness is hotly debated among scholars, who don't even agree on whether or not consciousness exists, much less of the qualities that would make up a proposed consciousness. However, the definition of consciousness that the average person has seemed to develop is that a conscious mind must have free will, and an awareness of its own existence in

the world around it. They also agree that humans seem to be the only minds that exists which is conscious. However, this apparent uniqueness gives rise to a sense of superiority, leading us to discount non-conscious minds. However, I believe that this sense of superiority is largely unfounded, and not necessary for a mind to be as developed and intelligent as our own. Furthermore, given the current technological trends, it's likely that a digital mind will be developed that, at least observably, demonstrates the existence of a consciousness. Digital minds, or "field-specific minds", which already surpass the best humans at tasks in specific fields, like GPT-2 for replicating language and Watson for playing Jeopardy, already exist, and as we further develop these minds we will be forced to reevaluate our internal hierarchy of minds and redefine what it means to be human.

Constructions of a Mind and Their Anthropocentric Leanings

In order to discuss the implications of conscious digital minds, I must first define the qualities that such a mind would take on, and the tasks that it must be able to accomplish. The most basic and earliest construction of an apparently conscious digital mind is one which can pass Turing's imitation game, or what is more colloquially referred to as the Turing Test. In Turing's seminal paper "Computing Machinery and Intelligence", he defines the rules for such a game: "It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'." (Turing 433). When replacing either person A or B with some

kind of computer program that responds to questions, it allows for testing whether or not a machine is indistinguishable from a human, at least through interrogation. Turing argues that a machine who can consistently fool an interrogator into thinking it is human must have a mind. However, Arlindo Oliveira complicates this idea of a mind in his book *The Digital Mind*, arguing that the existence of a digital mind is far more widespread than just those that can pass the Turing test. He proposes the " field-specific mind: a property of a system that enables the system to behave intelligently in a human-like way in some specific field" (Oliveira 221). This mind is much more prevalent in today's society; Oliveira argues that programs like AlphaGo and Deep Blue that excel at their respective games are necessarily minds due to the complex thinking required. Applying Oliveira's definition to the Turing test, no single field-specific mind is capable of passing the Turing Test, but a sum of minds that excel in different fields all working together would certainly be able to pass. This division of minds highlights the inherently human nature of field specific minds, which may not seem apparent considering the current state of technology and the abilities of modern field-specific minds..

Despite their usefulness in constructing a convincing human-esque machine, a major problem with the Turing Test and Oliveira's definition of field-specific minds are their inherently anthropocentric nature. Because they naturally involve comparing a mind to that of a human, it follows that we believe our own mind to be the highest form of intelligence, above all others. This belief follows from what Peter Singer describes in his paper "Speciesism and Moral Status", where he argues that humans incorrectly believe that they have a higher moral standing on account of our higher cognitive abilities. Singer cites philosopher Emmanuel Kant's beliefs for this idea, saying that "Kant's argument for why human beings are ends-in-themselves that

they are autonomous beings, which, in terms of Kantian philosophy, means that they are capable of reasoning" (Singer 7). Animals according to Kant are mere means to an end, which means that they have no autonomy. They simply exist on the grounds of survival, and are bound to their biological directives. On these grounds, he argues that our minds are superior to animals. However, we can apply Oliveira's concept of field-specific minds to animals. The biological goal of each animal is to reproduce and survive, and over time they have evolved to excel in that field. These qualities exemplify those of a field-specific mind, and thus I ought to include animals in the category of field-specific minds. This means that we can't simply base our superiority based on the fact that they have no mind. Instead, it is borne out of the animals minds not being "human-like". This superiority makes up the basis for a sort of "Human Exceptionalism" that directs how we compare our own mind to others. Barbara Finlay and Alan Workman define this term as being "a general proclivity to declare any difference between humans and their primate relatives a distinct human adaptation" (Finlay 1). The authors' usage of "proclivity" suggests that we are somehow hardwired towards believing that our differences somehow make us distinct and unique in an inherently valuable way. An important point is that this proclivity is born out of evolution, and not an objective superiority. Although this definition is a good starting point, I believe that Human Exceptionalism has more to do with specific qualities of our mind more than a general predisposition to identify differences between our mind and others.This term evokes the concept of American Exceptionalism, which Elizabeth Duqette discusses in her article "Re-thinking American Exceptionalism", saying that it is "promising the possibility of individual self-determination and prosperity, freedom from the rigid class constraints seen in other cultures, and the opportunity to worship and speak without fear of

reprisal." (Duqette 1). Essentially, Americans are able to succeed because they feel unfettered in their ability to think, and determine their own future. This idea can be extended to a species wide phenomenon, where humans feel "exceptional" because they have what we define as free will, consciousness, and are able to determine our future. Just like Americans feel free to define their career and life paths, we feel that our free will and consciousness gives us the ability to define our own immediate actions and totally control our future. On the other hand, so-called inferior minds like animals and digital programs have no will, and thus have no ability to control their future. If we were to strip back our human exceptionalism, both the Turing Test and field-specific minds fail to say anything valuable about the quality of a mind. They simply tell whether or not a mind is human in some fashion. Considering this, I think that we need to reevaluate the moral status of minds that typically wouldn't have any moral value considered to them.

Objections to the Raised Statuses of Non-Human Minds

A common objection to this raised moral status of digital minds is that their form of thinking is somehow inferior to ours, and shouldn't be considered to be equal to it. Turing details this objection, saying that "No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants." (Turing 446). His language is obviously evocative of human emotions, essentially saying that because they can't feel these abstract things that we do. In other words, digital minds cannot react to an outcome in the real world because they do not understand what it means to react in

such a way. Nicholas Carr raises a similar objection in his book *The Glass Cage, Automation and Us*, saying that "A computer, lacking any true understanding of the world, find the language of [a question] hopelessly ambiguous" (Carr 120). To him, because computers are unable to understand the world like a human can, they will never be able to accurately answer difficult questions with vague answers requiring legitimate thinking. Both of these objections raise the idea that digital minds are incapable of understanding abstract concepts that are either seemingly innate for humans, or are easily learned.

However, there are two main problems with this line of thinking. The first is that we have no methods by which to judge how a digital mind thinks. Our only well defined method to test the thought of a mind is the Turing test, and that only tests for human qualities in the mind. As Oliveira puts it in context to IBM Watson, the Jeopardy playing digital mind that was able to beat the best human players, saying that "Understanding is a gradual thing. It results from creating an internal model of the world that is relevant to the problem at hand. A system that has no understanding cannot answer, as accurately as Watson did, the complex questions posed in a game of Jeopardy. The model of the world in Watson's mind was enough to enable it to answer, accurately, a large majority of the questions." (Oliveira 229). Oliveira places Watson squarely in the category of field-specific mind. However, Watson can certainly, think, understand, but just doesn't exhibit human emotions. In many ways, it's thinking demonstrates a complex understanding of the world that many humans don't have, yet we would consider it lesser of a mind because of it. This links to the other problem with our belief in human superiority: the fact that we value human feelings, emotions, and unique creations over all else. GPT-2, a language generation AI created by OpenAI " generates synthetic text samples in response to the model

being primed with an arbitrary input. The model is chameleon-like—it adapts to the style and content of the conditioning text" (Radford). According to Carr, this AI does not understand the world, and it is "locked in its algorithms" (Carr 120), and even Radford opts to compare it to a chameleon, something we would consider to be a lesser mind. However, the generated examples from GTP-2 show a surprising grasp of the world around it. Given various prompts, it seems to understand that linguists study language, the University of La Paz is near the Andes mountains, and that a professor at La Paz university would most likely have a Spanish name. These connections in knowledge that we often take for granted require a vast understanding in the world, and I believe it isn't fair to say that GPT-2 does not possess such an understanding, especially considering the quality and breadth of it's writing. Still, many would scoff at the idea of it being similar to a human mind. Because of our human exceptionalism, we discount any other methods by which thinking occurs. However, both Watson and GPT-2 exhibit qualities that mirror human ones, often times surpassing our own abilities to think, with even further examples newly appearing of artificial intelligences that are able to interpret emotion. IEEE scholars created a digital mind with the ability "of making inferences regarding why a person (or an artificial agent) is in an emotional state using knowledge about emotions" (Rui 1), which informs decision making. As noted by the mind's use of inferences, and a databank of knowledge emotions, it's probable that the mind doesn't understand emotions in the same way Turing describes it. However, this objection is difficult to quantify; while I can certainly claim that my emotions are sophisticated, I have no method by which to evaluate the emotions of other humans in much the same way that I can't validate the emotions of a machine. We tend to assume that other humans have complicated emotions without evidence, yet we don't extend this assumption

to digital minds. I think this double standard is reason enough to disqualify Turing's raised objection as a requirement for consciousness. Similarly, Carr's objection fails to consider the fact that often times ambiguous questions stump humans just as much as they can stump digital minds. For example, non-English speakers can easily get confused by ambiguous sentences, leading to confusing translations reminiscent of a Google Translate query gone wrong. Native speakers of a language have years of cultural inclinations into what they should expect in a sentence, and that kind of large data crunching that humans naturally do in order to process language seems like a task perfectly fit for a computer to do. Saying that certain sentences are "hopelessly ambiguous" for a digital mind seems like a vast reduction of their true capabilities.

Conclusion

Given that, currently, we seem to be alone in our ability to form a consciousness, it may seem fair to declare a conscious mind itself as inherently valuable, and exceptional above all other minds. However, according to my definition of Human Exceptionalism, this is generated from our proclivity to value existential awareness and free will above all other traits of a mind. However, as shown by the trends of technology and existing digital minds, we must reconsider the functionally important values of a mind, and deconstruct the pedestal that we have placed our own minds upon. Of course, it's possible that a mind that truly matches the sophistication and development of our own never materializes, which would instead prove the inverse of my argument: that consciousness, and specifically the flavor present in humans, is inherently valuable and vital to a successful mind. However, considering the current trends of

technological innovation, I believe that a creation of a truly intelligent digital mind is only an inevitability.

When this happens, we will be forced to face the repercussions of such a mind. If we were to take the Turing Test to the extreme, and imagine a digital mind that could so perfectly mimic a human in every way, both physically and mentally. What does this say about the nature of being human? Essentially, you would have a human mind implanted inside a non-human body. One can imagine a similar routine, where instead of artificially creating a human-like mind, you transfer a human mind from one's natural body to a robotic one. These two robots were created in different ways, but have functionally similar brains. Which, if either of them, should be considered human? I think that these human-like minds should be considered human, requiring a total overhaul of our definition of humanity that cleaves it from its original species-based origins. The implications of this are vast: To examine one, we can point to the many arguments used by animal rights activists to assign some moral status to animals due to their ability to suffer. If we give machines the human status, can these arguments be extended down to lesser-minded machines? Their suffering may not have an analogue to human or animal suffering, but will most likely have some relation to the suffering of a human-like mind. Certainly we can't assign a moral status to all machines, and treat them with the same respect we give animals. This proverbial line in the sand must be hashed out, as it would have a large impact on almost every modern aspect of life. Ultimately, time will tell the story of whether machine rights are upheld or not, but I believe that we should begin making preparations to tackle these problems before we are forced to.

Works Cited

Carr, Nicholas. *The Glass Cage - Automation and Us*. WW NORTON & CO, 2015.

Duquette, Elizabeth. "Re-thinking American Exceptionalism." Literature Compass, vol. 10, no. 6, June 2013, pp. 473–82, doi:10.1111/lic3.12067.

Finlay, Barbara L, and Workman, Alan D. "Human Exceptionalism." Trends in Cognitive Sciences, vol. 17, no. 5, Elsevier Ltd, May 2013, pp. 199–201, doi:10.1016/j.tics.2013.03.001.

Oliveira, Arlindo"Brains, Minds, and Machines." The Digital Mind: How Science Is Redefining Humanity, The MIT Press, 2017.

Radford, Alec. "Better Language Models and Their Implications." *OpenAI*, OpenAI, 6 Nov. 2019, openai.com/blog/better-language-models/.

Rui Qiao, et al. "An Emotion Understanding System." 2014 2nd International Conference on Systems and Informatics (ICSAI 2014), vol. 461, no. Advances in Bionic Engineering, IEEE, 2014, pp. 264–70, doi:10.4028/www.scientific.net/AMM.461.85.

Singer, Peter. "SPECIESISM AND MORAL STATUS." *Metaphilosophy*, vol. 40, no. 3-4, Blackwell Publishing Ltd, July 2009, pp. 567–81, doi:10.1111/j.1467-9973.2009.01608.x.

Turing, A. M. "Computing Machinery and Intelligence." Mind, vol. 59, no. 236, Thomas Nelson and Son, Ltd., Oct. 1950, pp. 433–60.