"Artificial" Intelligence and Consciousness Through the Lens of Pragmatism

Technology is advancing at amazing rates, making our lives easier and creating tools that allow us to see and control the universe we live in in ways we could have never expected. One product of our technological advances is artificial intelligence, AI, which is a piece of software run on a computer with the intent of mimicking human consciousness and thought. Why we started on this path to creating a mind out of silicon is irrelevant, what matters is that we are reaching a point of no return that we are not prepared for. When the first true AI, that is to say, software with the ability to think and be conscious like a human, is created, we will have to decide how to, or if we even should, give it rights and personhood. Many philosophers have argued and will argue that these questions are irrelevant because machines will never be able to match our thinking, computers will only ever amount to simulations so we should not worry ourselves with being kind to them or thinking of their "feelings." The reason people argue against AI's ability to be conscious is an anthropocentric bias: they believe that humans have something special that computers just don't. This brings up a good question, what "special" thing do humans have that make us conscious? Going further, what is consciousness and what are thoughts? Consciousness is a term in our language game that has been only applied to humans, however, what we call consciousness is the ability to conceptualize sense data and participate in our language game. If any object expresses these qualities we must bestow the label "conscious" on to said object. The reason for our apprehension towards this view of consciousness is our deep-rooted anthropocentric bias.

Arguments Presented Against True AI:

This seems to be a straightforward enough definition, however, there are some who argue that computers will never be able to be conscious. I will focus on issues raised by John Searle because a) it sums up well the objections most philosophers have and b) it shows the flaw in the way of thinking of philosophers agreeing with Searle.

The main argument against true AI is best summed up by the Chinese room thought experiment. I will go through a quick summary. Let's say you have a man in a room, the room has a bunch of tiles with strange symbols on them, a window for tiles to come in and go out, and a book to tell the man which tile he should put out based on the tile that came in. Let us also say that this man gets so good at his job that his responses are instantaneous and he no longer needs the book. Unbeknownst to this man, the tiles he had were characters in Mandarin and he was answering questions in Mandarin. To native Mandarin speakers who are asking these questions, the man in the box appears to be a native speaker, however, the man in the box does not truly know Mandarin (Searle, "Minds"). The point of this thought experiment is to show that even though the man in the box appeared to know Mandarin, he did not actually understand Mandarin. Searle believes this thought experiment proves that intentionality is important when considering what has the capacity for understanding, he also believes this experiment shows what many philosophers argue: you can not get to a place of semantics from purely syntactic operations. Philosophers argue that human consciousness does not follow a certain set of syntactic rules but instead is something more "reactive" (Talks "John Searle"), therefore our consciousness cannot be mimicked by a machine that thinks by handling strings of ones and zeros. Searle has even gone as far as to say that if we get a running simulation of a human brain on a computer, this simulation will not be conscious either. Even though the computer is behaving exactly as a

human brain would because it is simulating an exact copy of the human brain, all it is is ones and zeros, so the computer has no consciousness (Talks "John Searle"). Some have supported this view by saying things like "A simulation of a rainstorm is not wet." Searle and many of his allies on this subject are obsessed with the idea of semantics and syntactics. They insist that consciousness is this complicated process that computers are simply incapable of running, and that the only thing capable of creating these conscious thoughts is organic matter. The only reason people make the argument that we are constructed from a different type of matter is that they believe that organic matter is for some reason superior to silicon, and this is a dangerous, chauvinistic view that I will discuss later. It should not matter what the material is made out of, or how the material completes the calculations needed to appear conscious. What matters is that the machine appears conscious for all intents and purposes and is able to interact with us. Searle himself, in the TED talk, said that consciousness is to our minds like liquidity is to a bunch of H_2O molecules. Liquidity isn't an extra substance created by H_2O, you can not separate liquidity from the individual molecules, instead, liquidity is the state in which the H_2O molecules lie. Similarly, no single neuron in our brain is conscious on its own, but put enough of them together in the right manner and consciousness appears. It can be said that consciousness, in this case, is also a state of conditions created by the grouping of the matter. Searle proves to himself here that consciousness is separate from matter and that the matter that makes up the system is not the only formation of matter that can achieve said system. He ignores this discovery and returns right back to arguing that computers will never understand the thoughts they are having like we do, simply because they are made of silicon. The arguments that shoot down true AI based solely on an issue with the material they're made out of are a sign of a way of thinking that places humans

above all others, and that begins with the idea that humans are special and tries to defend this belief. In the following sections, I will try to show how the term "consciousness" is not something that depends on the material it is presented in. I will be taking a dualist stance on it that I would like to believe many pragmatists would agree with. Instead of arguing for a sort of soul, I will be arguing that consciousness is not material but rather behavioral.

There are arguments that humans have a "soul" or "something special" that computers and animals do not. This is an important point because if you are to say animals are conscious, you will also have to say computers are, but I digress. To the people who make these arguments I say, find me that something special, and explain how it disproves my theories, until then, I will not address these arguments further. There are also arguments which Alan Turing calls "head in the sand" arguments (Turing 9). All that these arguments amount to are "if we are to grant computers consciousness, then bad things will happen." This includes arguments that say freeing computers will cause factories and society to collapse. I will answer all of these concerns at once, these objections are irrelevant. Just because the outcomes of a conclusion will be negative does not mean that the conclusion is any less true. You could say that that was an argument used against slavery in the past.

The Language Game:

One major flaw with the way we have been conducting all of our philosophies, not just the philosophy of mind, is that we believe the words we say have meaning separate from us, or to put it in better terms, we fail to recognize the language game. The language game is a term used by Ludwig Wittgenstein, and more recently Richard Rorty, to convey the notion that none of the knowledge we contain can be said to be statements about facts of nature, but rather statements of

what our peers agree on. If we are to accept the term "language game" to refer to how we communicate with each other, there is no clearly visible reason as to why our language game can not allow us to make statements about facts of the world. Where the issue arises with the language game is that our words do not have any intrinsic meaning, they do not have any meaning a priori to humans using them that we discovered (Rorty 175). An example of how our language game fails to connect with nature is the sentence "All dogs are animals." This sentence appears to be a fact, because by definition a dog is an animal. However, in order for this to be a fact it would require the terms "dogs" and "animals" to be categories defined in nature, outside of our language game. Unfortunately, they are not. Terms like "dog," "animal," and any other distinction we use to describe anything are created by humans in order to convey ideas amongst their peers, and even still, the understanding of what these words mean varies from person to person. In order for the terms in our language game to be statements of what is in nature some god-like deity, with all the knowledge of the universe contained in their mind, would have to descend from the heavens and tell us that our terms are correct. Since this scenario has yet to happen, there is no way we have access to the privileged knowledge needed to say our statements are statements of fact. What is there then? Since our language game keeps us from making statements about nature separate from human experience, how do we go about knowing things and discussing ideas? Rorty's solution to this is what he calls epistemological behaviorism, which frames the study of knowledge as the study of how people interact (Rorty 175). Rorty's epistemological behaviorism changes how we view the statement "S knows P." While more traditional, foundationalist, philosophers will say this statement references the relationship between S and P, Rorty says that we can view this statement as a claim about S's

report of P to their peers (Rorty 175). Epistemological behaviorism defined by Rorty is the method we already use for determining truths and living in the world, Rorty would call this philosophy pragmatism, but he did not wish for it to be related to the term. Rorty continues on to assert that there is no such thing as knowledge that is non-propositional, and no such thing as a justification of a belief that is not a relation of propositions (Rorty 183). This way of putting our methods of knowing things and justifying our beliefs is not outrageous, in fact, Hume, an earlier philosopher, touched on how all of our thoughts can be said to be impressions or a relation of impressions. Hume's observation came after he brought up an issue with proof by induction. Hume realized that we have no reason to believe that what has happened in situations will continue to happen in like situations (Hume 15-16). Hume's solution to this was to come up with a pragmatic way to describe how we think, in which we do not need to know something to be a fact in order to treat it as a fact. Instead, we slowly begin to relate experiences we have until we have a belief about the world, then we test that belief against experience and begin to create a lexicon of facts.In this case, facts is taken to mean a piece of knowledge that we believe to be a fact, in our own minds (Hume 18). At this point it may seem like all knowledge is worthless since it is simply a part of the larger language game. However, only the philosopher wants something more than epistemic behaviorism because only they are afraid of the alternative. All non-philosophers have always used this method to justify everything in their lives without worry (Rorty 181).

The issue with "consciousness":

What I have argued in the previous paragraph may seem pointless. I first state that the terms we use to describe the world cannot actually reference what is in the world.Then I go on to

say that this does not matter and we have been able to deal with this for all of human history. The significance of the language game is that none of our terms or ideas have any discovered meaning, "dogness" is not something we discover. It is something we give to animals who appear to be dogs for all intents and purposes. Similarly, "consciousness" is not some mystical quality that we were looking for, and once we discovered it, the word "consciousness" manifested itself in our minds.It is a term we have created in order to convey to our peers what we think of when we think "conscious." What this means is that there is something special that makes humans conscious, and that special thing is there are other humans who will look at us and agree we are conscious. The term "conscious" has no meaning outside of our language game and is not a discoverable physical trait of something. Instead, what consciousness is is the ability to conceptualize what you experience and see, and report it to your peers in an acceptable manner.

   Children are said to be conscious by most people, however, a newborn cannot be said to "know" the color red differently from any photoelectric cell. A child that has not started speaking yet "knows" what red is in the sense that it can behave discriminately towards the color red, an action which all digital devices are able to do (Rorty 183). Rorty argues that the reason we are so convinced that a baby is conscious, despite its behavior being at a similar level to that of a computer, is because in a baby we see ourselves and the potentiality of conversation. We understand that a human baby will one day be conscious, so we feel comfortable granting this conscious to it since birth. This is a mistake because we are assuming the premise that babies are conscious, not asking the question, are babies conscious?

If we are to take consciousness to be as I have defined it in this essay, then the answer is no, babies are not conscious. I believe Rorty would take my side in this as he argues in his book that consciousness being granted to a child who begins to learn to speak and communicate is analogous to rights being granted to a teenager when they become 18 (Rorty 187). Turning 18 does not change something inside of you that turns you into someone deserving of these rights, instead, turning 18 changes the way society looks at you. While many people believe society grants us rights because we deserve them, this example demonstrates the converse is true; we deserve rights because society says that we do. It is because of this that Rorty feels comfortable saying babies learn consciousness, it is because consciousness is not some switch inside of a creature but the way our society views and interacts with that creature. Still people will argue that babies are clearly conscious, they are moving around and interacting with their environment, clearly, that is consciousness! I will concede that a baby appears conscious, and if someone wants to define consciousness to be "the ability to react to their surroundings," then I will agree that consciousness should be defined that way. Sellars addresses this response by accepting the definition of consciousness, but also demanding a new term to describe when someone has the ability to conceptualize data and argue about that one concept's relation to other concepts held (Rorty 182-184). A baby can know what pain is like, that is to say, experience the sensation we call pain, however it does not know what type of thing pain is. A newborn may know to avoid things that cause the sensation of pain, but this newborn does not know that pain is what most people call a negative sensation. In fact, the newborn has no concept of the terms "sensations" or "negative". It is the ability to conceptualize data and make arguments out of them and create relations that Sellars, Rorty, and I want to call consciousness. In the given example with the

baby, the child will one day learn how to convey what hurts it and what it enjoys, it will be able to understand the world it is experiencing in the way we do and demonstrate this to us. When this occurs the child can be said to be conscious in the same sense that we, as adults, are.

The Issue With Anthropocentrism:

I have spoken a lot about how people have a chauvinistic tendency to say humans are in some way superior, or special, but have had yet to talk about why this is bad. To start off, believing humans are somehow special leads us to just accept without question certain premises. An anthropocentric bias also causes us to mix up traits necessary for consciousness with traits that are merely how humans exhibit consciousness. We are willing to say that koalas are conscious and have feelings similar to ours, even though koalas are unable to fully grasp what their food is. If you put a koala in a room filled with eucalyptus leaves, its diet, the koala would starve because it does not know what eucalyptus leaves are if they are off of the branch. The reason we so willingly grant creatures like koalas consciousness is that we see a part of ourselves in them. One last concern is that people are resistant to philosophical theories that makes them less than the idea of themselves they have in their minds. If up until reading my paper, someone believed humans had something special that made them more than a set of rules being run on a complex machine, the views I have presented them must sound very upsetting. Someone holding an anthropocentric bias will work hard to defend consciousness as something special and will bend their arguments to force compliance with this closely held belief because in the theories I have been discussing, consciousness is nothing special at all. I am not saying consciousness is worthless or something we should not be thankful to have, I am pointing out that what consciousness is is not as special or unique as a god given soul. The hangup with a computer

program being solely syntactic operations is, I argue, a issue of anthropocentric bias as well. We do not feel comfortable saying our consciousness can be expressed in a list of syntactic rules, so we instead say that, as a premise, consciousness cannot be expressed in a list of syntactic rules.

Where Computers Come In:

Long ago, Alan Turing began thinking about if a computer could become conscious. His solution to this question was related to the imitation game, the gist of this experiment is that a computer tries to trick a human judge into believing it is human based on answers to questions (Turing). Turing came up with a very pragmatic solution early on, his philosophy revolved around the idea that if it behaves like a conscious human, it is. Many people have criticized Turing for creating a test that puts the bar for human consciousness too low and for his apparent disregard for intentionality when it comes to thought. I agree that his initial idea for a test is too easy to beat, and some software has passed the official Turing test, but he had the right idea when he came up with it. My suggestion is that instead of a computer having to pass a written test for the Turing test, the program has to be able to live and interact with society. If a computer can actively behave like a person, then it is one for all intents and purposes. Hopefully, by this point, I have shown that consciousness is not a discoverable physical quality, so it should not be too outrageous for me to say that the key to calling something a person is to see if it can behave like one.

One unique issue that has been brought up to me is that of self consciousness, knowing that you are a thing that exists in the world and being able to ask questions about this. Self consciousness is a trait that humans have but is also lacking in many animals. Does this mean that self consciousness is the defining characteristic of consciousness? Well, no. We need to

remember that self consciousness is merely another term in the language game, so there are not necessarily component pieces that make up consciousness. That being said, self consciousness is important in the sense that it allows the creature to understand how to interact in the world. The lack of self consciousness in computers currently does not disqualify them from being able to have it in the future. Self consciousness appears to be an extension of the ability of an individual to conceptualize the body they are inhabiting and the thoughts they are having, in this way self consciousness is a result of how consciousness is defined in this paper. If we see self consciousness as the manifestation of the desire to understand what one is, we can see that this can be programmed into computers (Jenkins 4). The inclusion of self consciousness adds another important features to computers, the ability to participate in our moral society. In an earlier essay, I made the point that if a creature is able to communicate its desires with us, able to understand our desires, and is able to understand the consequences of its actions, it can make moral claims against us and have claims made against it (Namiranian).

So What?:

One question that I have failed to answer in this essay is why does any of this matter? That is a very important question to think about, what difference does it make whether or not computers can be said to be conscious? Why I chose to argue that computers can be conscious is because this way it can be argued that they deserve rights and personhood. If a computer is able to participate in our moral community, that means that its suffering is the same as ours. Since our current society relies heavily on computers in a large portion of their lives, if we create conscious computers we are creating an entire slave race. Even if they do not have the same level of consciousness that we do, it is naive to think that we can completely ignore discussing their

rights because they are not organic. What is important going forward is that we recognize our natural tendencies to think humans are something special, not only because it will give consideration to the rights of other potentially conscious beings, but because it will allow us to conduct philosophy without clouded judgment.

Bibliography

Dennett, Daniel C. "Consciousness Explained." zodml.org : Little, Brown and Company, 1991. https://zodml.org/sites/default/files/ %5BDaniel_C._Dennett%5D_Consciousness_Explained.pdf

Hume, David (1977), "An Enquiry Concerning Human Understanding", Eric Steinberg (ed),. Indianapolis: Hackett Publishing Company.

Jenkins, Scott. "Self-Consciousness in the Phenomenology." In The Oxford Handbook of Hegel.: Oxford University Press,, 2017-08-31. http://www.oxfordhandbooks.com/view/ 10.1093/oxfordhb/9780199355228.001.0001/oxfordhb-9780199355228-e-5.

Namiranian, Pedram. "Does My Toaster Get Rights?", December 14, 2017.

Rorty, Richard. "Philosophy and the Mirror of Nature by Richard M. Rorty." Goodreads, Goodreads, 21 Jan. 1981, https://pages.uoregon.edu/koopman/courses_readings/rorty/ rorty_PMN_full.pdf

Searle, John R. "Minds, Brains, and Programs." In Philosophy of Mind: A Guide and Anthology. Ed. John Heil. New York: Oxford University Press, 2011. 235–52.

Talks at Google. "John Searle: "Consciousness in Artificial Intelligence" | Talks at Google." Filmed. YouTube video, 1:10:37. Posted [December 2015]. https:// www.youtube.com/watch?v=rHKwIYsPXLg&t=632s

Talks at Google. "Ned Block: "Why AI Approaches to Cognition Won't Work for Consciousness" | Talks at Google." Filmed. YouTube video, 1:03:10. Posted [June 2017]. https://www.youtube.com/watch?v=6lHHxcxurhQ&t=1596s

TEDxCERN. "John Searle: Our shared condition - consciousness." Filmed. TED video,

14:55. Posted [May 2013]. https://www.ted.com/talks/

john_searle_our_shared_condition_consciousness?language=en

Turing, Alan M. "Computing Machinery and Intelligence." In Philosophy of Mind: A

Guide and Anthology. Ed. John Heil. New York: Oxford University Press, 2011. 212–34.