|  |
| --- |
| Machine Consciousness |
| Majok RingWRI 100March 31, 2016 |

# Abstract

Within the sciences, widely accepted theories of consciousness such as Theory of Mind focus primarily on an anthropomorphic viewpoints. Theory of Mind follows two conditions to determine consciousness. 1. The ability to perceive your own consciousness and 2. The ability to perceive the consciousness of another. Although this may be helpful in understanding our own consciousness, this viewpoint becomes limiting in that we are unsure if it can be universally applied to an area such as machines and Artificial intelligence. Utilizing prominent studies, this paper will go over the popular definitions of consciousness such as Theory of Mind and will bring to light questions concerning how this theory melds with computationalism in machines and will attempt to analyze any present setbacks in logic.

# Introduction

There have been many prevailing definitions and conclusions on consciousness available throughout time. It can be observed that the most accepted theories are often ambiguous in nature and allow us to delve deeper into etymology in order to find the close distinctions. Our current societal definition of consciousness is “the fact of awareness by the mind of itself and the world”. Although the definitions of consciousness continue to change over time, the key aspect of awareness remains constant. This change of definition can be highlighted through ancient philosophers such as Plato, Aristotle, Descartes, and even later figures such as John Locke. Descartes’s claims conflict with previous theories as he states that the mind is a non-physical substance and defines consciousness as thinking very broadly to include every kind of mental state available. This is highlighted in his work, the Principles of Philosophy (1640), in which he states “By the word ‘thought’ (‘pensée’) I understand all that of which we are conscious as operating in us.” This aspect of ambiguity has also been probed by Locke in his written work, An Essay on Human Understanding (1688). Here, Locke states “I do not say there is no soul in man because he is not sensible of it in his sleep. But I do say he cannot think at any time, waking or sleeping, without being sensible of it. Our being sensible of it is not necessary to anything but our thoughts, and to them it is and to them it always will be necessary.” As time progresses, we see our theories on consciousness continuously become more defined, yet may still be too vaguely developed to apply universally to all entities either human or machine. With the advent of the 20th century, new and popular theories such as Theory of Mind, computationalism and Emergence theorem develop and allow us to pose new questions about how we should define consciousness. With Theory of Mind (ToM), we are able to describe consciousness through the ability to be metacognicent of your own existence and also the ability to attribute mental states to others. Essentially, this means that if you can understand that you are in control of yourself and are completely independent and have the ability to observe this attribute in other beings, then you yourself must be conscious. This contrasts directly with computationalism, the understanding that our behavior is the result of numbers of complex computations within our mind and is not reliant on self-perception or cognition. This theory branches off and may allow us to broaden the definition of consciousness enough to apply to machines and artificial intelligence as it relies heavily on computation for behavior. Another important theory which remains is Emergence Theorem. This theory states that as time progresses, through small events and interactions produces dramatic changes towards an unforeseen outcome. Essentially, this means that since many systems may be very complex, we are unable to account for all instances/effects and may end up receiving unexpected results. With regards to artificial intelligence, by continuing small amounts of progress and advancements, we may see machine consciousness at a specific point in time emerge suddenly.

# Computation and Emergentism

In order to see if the current view of consciousness within the scientific community holds against scrutiny, it is important to be able to view the aspect from several different viewpoints. This is argued within American cognitive psychologist Arthur Reber’s “Caterpillars and consciousness” (Reber, 1997). Now the idea forming is that our understanding of consciousness is based too much on an anthropomorphic viewpoint and can lead to misleading data and ultimately a narrow definition of consciousness which may allow scientists to come to erroneous conclusions. Here, scientists may try to apply the same logic that is used for humans and define consciousness in a way that is applicable to only organic life. Within Reber’s essay, he describes the act of observing a caterpillar and understanding the fundamental aspects of its motion and behavior. It was found that they were mechanical, routinized, and repetitive, similar to a machine. Reber clearly explains that anthropomorphism can be an entertaining concept and has greater prevalence in theories as it allows for easier study as we are able to relate to concepts more quickly and apply our own understanding through personal experience. This highlights our need to relate behaviors between ourselves and other entities and can be the reason why we have a limited view of consciousness as we are using only ourselves for the criteria. Now, if scientists base AI off of our own likeness and attempt to relay our language, behaviors, and thinking processes to a machine, will that be enough to define the machine as conscious if that is what our criteria is based off of? Although it should be enough based off of accepted definition of computationalism, problems in defining it arise as true intelligence requires intentionality, self-reflection, decision making, problem-solving skills, creativity and also rationality.

AI may have direct issues with creativity and self-reflection as it is difficult to prove artistic intention and provide a distinction between genuine creation and simple shuffling/randomization of patterns. With regards to software development incorporating AI, it is simple to program a wide variety of embedded responses, but these are not valid as actual behaviors, but simply put, are a series of controlled reactions. It is for this reason that AI has difficulty with creativity. Another important aspect of emergentism is present through our tinkering with AI. By continuously looking to advance and integrate more anthropomorphic features onto our machines, we are unknowingly going to reach a sudden transition stage where we find ourselves with machines that are actually conscious.

# C:\Users\Majok\Desktop\3dae4a000d4da699014800205f38d629.jpgCreativity

With regards to the aspect of creativity, we generally cannot define machines as creative as their responses are pre-programmed and merely react to stimulus. Take for example conversations with chatbox AI. Within these conversations, the AI is not actually conversing, but recognizes certain keywords and can alternate its response based upon those keywords. Essentially, it is limited to what it has had programmed into it and cannot efficiently learn or create new things on its own.

Figure 1

Created using Google’s Deep Dream

Image by Kevin Mark Rabida

July, 8th, 2015

In Figure 1, we will analyze this photo which was created using Google’s Deep dream generator. DeepDream is Google’s computer vision program which utilizes a convolutional neural network in order to develop these images. These convolutional neural network is an artificial neural network which is organized to mimic an animal’s visual cortex and respond to stimuli in a similar fashion. In order to develop this network, Google utilized a bottom-up approach by attempting to develop a system in which mimics features and organization of the brain in order to function effectively. Within the runtime of this software, an algorithm is ran in which enhances pareidolia (mind perceives patterns when they don’t exist) within the image creating a hallucinogenic cast on the processed image. This is akin to seeing faces or illusions in the dark as the human mind attempts to reorganize shapes and shadows into familiar objects/images. This is an interesting concept as it relates back to one of the core concepts of Theory of Mind: interpreting others as conscious. Within this process, the machine generally looks through the image and finds what could be a facial feature and continues to enhance and make the features more apparent with each repetition. After the final process we are able to see a new image with many differences as buildings and vehicles may be turned to slugs and birds and any eyelike hole or crevice may be turned into eyes and later turned into a dogs face. This DeepDream simulator seems to generally gravitate towards seeing dogs within every image and converting the subjects into dogs. Now, although the produced images look artistic and creative, we have to look deeper into the context of how it is completed. This process being done by the machine is not done with any creativity in mind. It is rather simply an algorithm that looks for areas of noise which matches a certain pattern and applies certain effects from its database onto the area. This concludes that it is not actually creating purposeful art, but is merely undergoing an algorithmic process.

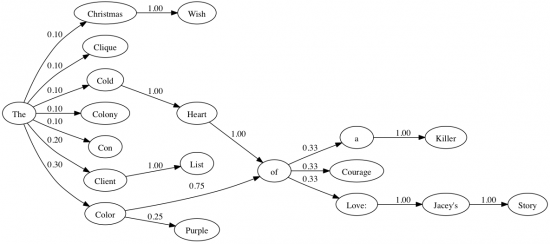


Figure 2

Markov Chains: Generate Movie Titles

Image by Jason Bury

In figure 2, we are presented with an example of a Markov chain that generates movie titles based off of data collected from servers. Markov chains are defined as a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. Essentially, this is one of the many ways in which a machine is capable of thinking. Each process must be calculated by the process before that and relies heavily on probability and making inferences from available data. This experiment serves the purpose of testing the results probabilistic thinking of machines and also the structure in which this thinking is formed and providing a comparison to human results. This procedure for this experiment required a computer script to scrape 1,968 titles from a movie channel website (Lifetime Movies) and implemented a Markov chain to model the likelihood of and ordering of certain words. Take for example the word “The”, in this experiment there is a 10% chance of any word being selected next, but with words such as “Christmas”, there is a 100% chance of the word “Wish” being chosen afterwards or also if the word “Cold” corresponds with a 100% change of choosing “Heart” as the next word. This is based off of data and patterns that are recognized from the machine. After this process is completed and the probabilities are calculated, it is possible to create new titles from the ones available based off of predictive computing. Examples of the generated titles include “A Crime of Love and Innocent Victims”, “Dr. Quinn: The Perfect Alibi”, and “A Nightmare at 17”. In terms of success, the titles created by the AI are almost impossible to distinguish as auto generated titles and are similar to the thought process that a human would undergo when making titles. A human would look to popular movie titles and genres for inspiration and would be influenced by those titles as well as a machine.

# An Ethical Machine

Generally, we find that science focuses on asking questions such as to the possibility of a conscious machine and bases clues off of our current understanding. Rarely do we ask if we should develop conscious machines and the address the ethics behind this process. Within the article, Ethics, Regulation, and the new Artificial intelligence, Part II, written by Perri, we can see the concerns behind conscious machines and the impact they could have on public policy and even military applications (Perri, 2001). Here, we can see the discussion behind public policy as AI may have an increased impact on human decision making and a greater degree of autonomy. This is a concern as we have no way of currently discerning if a machine is fully conscious and able to accurately interpret all scenarios and reach ethical results. Take for example the trolley problem. This is a well-known thought experiment which has the scenario where a runaway trolley is barreling down the tracks and there are five people tied up, unable to move. Since the trolley is going to run over and kill those people, you have the ability to pull the lever and divert the trolley and save those people at the cost of a single person on the other side being killed. Now, most people if they have no familial connections to the victims may choose to save five people at the cost of the single person. In this case, we are able to do this by rationalizing our decision through beliefs such as the single person should not have been on the tracks and believe that the lives of five people outweighs the life of one. Now, a machine given the same scenario may choose the same outcome as it is a logical decision to save five instead of one. The machine in this case will not need to do any emotional rationalizing and will come to the decision immediately. Now, if we alter this scenario a little bit and state that five people are tied up, but if you pull the lever, the train will crash and be put out of commission for years. In one scenario, five will die, but in the other none will die. Now, although not many people would sacrifice a single human life for a machine or infrastructure, a machine in this case, may opt to let the five people die as their worth to society is not as great as the worth of the train. The train may take thousands of people to work per day and provides a very important utility towards the city in terms of transportation and assists the city’s GDP in terms of productivity. Now, since trains may be expensive and difficult to repair, the machine will easily take the five lives and spare the train in order to produce what it believes to be the most rational outcome. This scenario displays the concern that others may have with conscious machines and their expanding applications to public policy and military applications.

# Conclusions

Throughout this paper, I have analyzed the continuously changing definition of consciousness and have attempted to decide if it is universal with respect to machines. This was done using a comparison of the different available theories such as Theory of Mind, Computationalism, Emergentism, and also was assisted by analyzing different approaches such as bottom-up approach. Due to the fact that the definition of consciousness is continuously changing as time progresses and cannot keep up with progress and development in the sciences and also the various conflicts and lifespan of available definitions and theories, we cannot plausibly accept that our definition of consciousness via Theory of Mind, is applicable to machines and other non-human entities. Although there are many similarities in capabilities of producing and creating new and taking inspiration from what is available, the same motivations are not at play for humans and machines. This was explained through the examples of art as we define art as visual/auditory works created with a purpose in mind. With machines, that purpose was not self-inspired and often times was simply following a set of instructions or algorithms to produce the artwork. Although there is not currently a theory which is capable of accounting for all of the overlapping areas and lacks conflicting logic that does not mean that it is not possible for the future. This is explained via Emergentism as we continue to search for questions to the definition of consciousness and attempt to apply those principles further and further to machines/AI, we may end up reaching this scenario through unexpected means.

# References

Reber A. 1997. Caterpillars and Consciousness. Philosophical Psychology (10) [Internet]. [N/A, cited 2016 Mar 17] Vol.10. Issue. 4. Available from: http://www.tandfonline.com/doi/abs/10.1080/09515089708573232?journalCode=cphp20

Gualtiero P. 2010. The Resilience of Computationalism. *Philosophy of Science* 77, no. 5: 852-861. *Academic Search Complete*, EBSCO*host* (accessed March 17, 2016).

SEBASTIÁN, M. Á. (2016). Consciousness and Theory of Mind: a Common Theory?. Theoria, 31(1), 73-89. doi:10.1387/theoria.14091

Dournaee, Blake H. 2010. "Comments on “The Replication of the Hard Problem of Consciousness in AI and Bio-AI”." Minds & Machines 20, no. 2: 303-309. Academic Search Complete, EBSCOhost (accessed March 17, 2016).

Perri. 2001. "Ethics, regulation and the new artificial intelligence, part II: autonomy and liability." *Information, Communication & Society* 4, no. 3: 406-434. *Academic Search Complete*, EBSCO*host* (accessed March 17, 2016).