John Searle's Argument

Anne Reynolds

The topics of strong AI and weak AI are frequently fought over among the computer science community. Strong AI is defined as a machine that has sentience, consciousness, and a mind ("Weak AI", 2017). Whereas weak AI is a non-sentient machine that can only perform one specific task ("Weak AI", 2017). An example of weak AI would be Deep Blue. Deep Blue was created by IBM and it attempted to "learn" all of the possible moves in chess so that it would always know the optimal solution and win against opponents. While it utilized generative processing algorithms, it only focuses on performing one specific task: learning every optimal chess move. Currently, there have been no formally recognized examples of Strong AI, however this isn't to say that there haven't been valiant attempts. A recent attempt at Strong AI is IBM Watson. Watson collects a large database of knowledge to answer questions posed in natural language. It uses machine learning algorithms and parses for keywords and then decides based on probability. It is used currently commercially in many different domains of businesses ("Watson (computer)", 2017*)*. While IBM's Watson is useful, it is not an example of strong AI.

A critic of the existence of strong AI is a John Searle. John Searle is a renowned American philosopher who has used his concept of consciousness to deny the existence of such phenomenon. He proposed his rebuttal to strong AI in the form of the "Chinese Room". In this example, one has to image that they are in a room with a small slit, a book and a piece of paper. The participant has no knowledge of Chinese. Chinese characters are given through the slit and the participant must follow the directions written in English in the book. They then complete the task of constructing some string of chinese characters and put the paper through the slot. Outside of the room it may appear that the room understands Chinese. However, the person inside is just following instructions and actually has no comprehension of the Chinese.

Searle thus concludes that there is no possible way that a computer could ever understand English or Chinese even if the outputs are correct.

This argument was not met without critics many critiques. Schank claims that his AI understands the story and can provide answers to the questions. He also claims that his machine explains human cognition in a way that explains how we decipher and analyze stories and answer questions ("Minds, Brains, and Programs"*,* 2003). Searle is not convinced in the slightest that this is a sound argument and he uses the Chinese Room to argue his case. For example, if you're given a set of symbols in a language you do not understand and rules in a language you can understand, you may get the correct output, however there is no real understanding of the symbols or their actual meaning and relation to others in the language. You can be given a string of sentences in Chinese and be able to accurately manipulate the symbols according to rules in order to create valid sentences. However, that does not reflect your understanding of the language of Chinese. Searle said that he has the same input and output as a Chinese speaker, however you could still not understand the language that is relevant("Minds, Brains, and Programs"*,* 2003).

In his rebuttal, Searle uses the Turing Test to disprove this theory. He says that two machines could pass the Turing test, and fool Chinese speakers. However, only one machine could understand Chinese. The machine is simply manipulating symbols from a given input and presents the output without any understanding of outward perception or consequences of the manipulation ("Minds, Brains, and Programs"*,* 2003).

Searle also claims that in order to understand and define Strong AI and for it to be reputable, you have to distinguish clearly principles on how the mind works versus how non-mental systems work ("Minds, Brains, and Programs"*,* 2003). There has to be a definition that is precise and can be applied to all strong AI, otherwise we could attribute hurricanes to being mental beings and humans as being non-mental. The definition has to be consistent among all for it to survive as a theory.

Someone from Berkley argues that if you could simulate a machine to fire synapses in exactly the same way a Chinese person would, you couldn't refute it or else you would have to deny that a Chinese person understands Chinese ("Minds, Brains, and Programs", 2003). Searle replied with saying that the idea of strong AI is that we do not need to know how the brain works in order to understand how the mind works ("Minds, Brains, and Programs", 2003). If we had to know how the brain worked to do AI, it wouldn't be worth it. The person from Berkley's theory simulates the "wrong" things in the brain. By this he means that If it simulates the formal structure, it will not simulate what matters in the brain - its ability to produce intentional states ("Minds, Brains, and Programs", 2003).

To the argument that is a combination of the two prior arguments, Searle claims that it is insufficient to claim that a robot has mental capabilities just because it looks and behaves like a human ("Minds, Brains, and Programs", 2003). To the person at Yale who ascribes mental based on behavior he says that you can't judge the validity based on computational processes and their output because they can exist without a cognitive state ("Minds, Brains, and Programs", 2003). A cognitive state is necessary to claim that there is mental activity occurring .

In order to develop a machine that can truly be identified as being a strong AI, there must first be a universal definition that can apply to all machines in all situations. There also needs to be a universal understanding of consciousness and the degrees to which you can claim that there is conscious understanding. A machine must have general knowledge and not just perform one task. However, strong AI may never surface because behavior is not a sufficient form of judgement of intelligence.

References

(2003, July 14). *Minds, Brains, and Programs.* Retrieved from http://cogprints.org/7150/1/10.1.1.83.5248.pdf

(2017, April 25). *Weak AI.* Retrieved from https://en.wikipedia.org/wiki/Weak_AI

(2017, April 30) *Watson (Computer).* Retrieved from https://en.wikipedia.org/wiki/Watson_(computer)