

Chinese room  
thought experiment  
philosophy essay



**ASSIGN  
BUSTER**

The mind has been the center of philosophical debates for the longest of times. John Searle has attempted to explain understanding and the mind when in 1980 he created his famous Chinese Room thought experiment. However, he is not discussing the human mind like many philosophers do. Instead, he is looking into the minds of machines. Searle is looking into Artificial Intelligence and debating whether or not it can actually be comparable to human understanding. First I will give an overview of Searle's description of Artificial Intelligence. Second, I will explain the Chinese Thought experiment and its implications. Third, I will describe five of the most common responses to Searle's thought experiment. Finally, I will analyze the thought experiment and see what conclusions can be drawn. But first let us discuss Artificial Intelligence.

While the Chinese Room thought experiment was originally posed to counter the claims of Artificial Intelligence researchers, philosophy has also used it to look into the minds of others. It is a challenge to functionalism (mental states constituted solely by the role they play) and the computational theory of mind (the human mind is information processing system and that thinking is a form of computing) and is related to many others famous thought experiments.[1]In the Artificial Intelligence debate, Searle analyzed a position which he refers to as strong Artificial Intelligence. This position claims that a computer with the right program, with the right inputs and outputs would have the same mind as any human.[2]

However, there is a distinction that needs to be made between strong Artificial Intelligence and weak Artificial Intelligence. Searle writes that “ according to Strong Artificial Intelligence, the correct simulation really is a <https://assignbuster.com/chinese-room-thought-experiment-philosophy-essay/>

mind. According to Weak Artificial Intelligence, the correct simulation is a model of the mind.”[3]In other words, Strong Artificial Intelligence is equivalent to the mind while weak Artificial Intelligence just has the appearance of the mind.

John Searle in 1980 developed a thought experiment that is supposed to determine that Strong Artificial Intelligence is impossible. Imagine that engineers succeeded in creating a computer that understands Chinese. The computer “ reads” Chinese characters and, by following a computer program, “ writes” other Chinese characters. Searle imagines that this computer is so convincingly that it passes what is referred to as the Turing test, that is to say that the computer convinces a fluent Chinese speaker that it is also a fluent Chinese person. In other words, the fluent Chinese speaker is able to carry on a conversation with the computer without ever suspecting that it is a computer. In the second part of Searle’s thought experiment he imagines a similar scenario. However, instead of a computer, we have an English speaking person that has no knowledge of how to speak Chinese in an isolated room. However, this person does have a book with instructions similar to that of the computer program in which he is able to take the Chinese characters that he receives under the door use the book, write a response, and slide it buck under the door. Like the computer, he is also able to convince a native speaking Chinese person that he can speak Chinese fluently.

Searle claims that there is no substantial difference between the two cases.

Both the computer and the English speaking person are following a program/ instructions which simulate understanding. However, we would not claim <https://assignbuster.com/chinese-room-thought-experiment-philosophy-essay/>

that the English speaking person understands Chinese, and therefore, we must conclude that the computer also does not understand Chinese.

Searle argues that without understanding, we cannot be described as thinking, and because the computer doesn't understand, we cannot describe what the machine is doing as thinking. Therefore, Searle concludes that strong Artificial Intelligence is not possible.

Searle argues that without understanding, we cannot be described as thinking, and because the computer doesn't understand, we cannot describe what the machine is doing as thinking. Therefore, Searle concludes that strong Artificial Intelligence is not possible.

While this thought experiment has been around for a relatively short time, it has already received several objections. The most common response to the Chinese Room thought experiment is what is referred to as the system reply. In this response, we concede that the man in the room does not understand Chinese. However, the man is only a single part of the system just like the CPU is only a single part in the computer. Both the man and the CPU have help from other components like paper for the man or memory for the computer, and it is the system as a whole that understands Chinese. According to Ray Kurzweil, the man is comparable to the CPU and is therefore an implementer that is of no importance.[4]Kurzweil also agrees with the Turing Test and agrees that if the system shows that it can apparently understand Chinese, it would have to understand Chinese. He goes on to claim that Searle is contradicting himself saying that the computer speaks Chinese but does not understand Chinese.

However, Searle responds to the claim that understanding is the sum of physical objects like the book and papers. Searle simplifies the thought experiment by removing physical objects. Imagine if the man memorizes the book of instructions, and keeps track of all the information he otherwise would have written on paper in his mind. Then, according to Searle, the whole system is just the man and Searle argues that the man does not understand Chinese and therefore the system does not understand Chinese, and even though the man appears to understand Chinese, it proves nothing. [5]

However, proponents of the system response claim that this alternative still does not dismiss the claim that it is the system that matters. These people claim that the book being in the man's head has allowed the man to have, in a sense, two minds.

The second response to the Chinese Room thought experiment is often referred to as the Virtual Mind response. This response is very similar to the first response in that those who agree with the Virtual Mind response agree that the man does not understand Chinese by following the instructions. However, unlike the systems response, the Virtual Mind response believes that when you run the program in the computer or follow a set of instructions for the man it creates a virtual mind.

The word "virtual" is a computer term that is used to describe an object which seems to exist "in" a computer. Objects such as files and folders appear to exist only because software is making it appear to exist.

It is argued therefor that while the computer might not understand Chinese, a virtual mind inside the computer could understand and use Chinese.

However, Searle objects to this response saying that this virtual mind is Weak Artificial Intelligence i. e. a simulation. Searle writes: “ No one supposes that computer simulations of a five-alarm fire will burn the neighborhood down or that a computer simulation of a rainstorm will leave us all drenched.”[6]This objection is not shared by everyone; Nicholas Fearn suggests that some simulations are as good as the real thing. “ When we call up the pocket calculator function on a desktop computer, the image of a pocket calculator appears on the screen. We don’t complain that ‘ it isn’t really a calculator’, because the physical attributes of the device do not matter.”[7]

The third response to Searle’s thought experiment agrees with the whole thought experiment. The Robot response agrees with the Chinese room response argues that with some modifications the computer would begin to understand. Suppose that the computer is built into a robot that could walk around and interact with its environment. By placing the computer inside a robot it would allow a “ causal connection” between the symbols in the programing and the objects it represent much like a child is able to draw connections between new words and new objects. According to Hans Moravec, “ If we could graft a robot to a reasoning program, we wouldn’t need a person to provide the meaning anymore: it would come from the physical world.”[8]

Like the other responses, Searle has an objection to the robot response. According to Searle, the sensors just provide additional input to the computer, and as a response just additional data, and not understanding. This can be echoed for the man in the room. " Suppose the man in the Chinese Room receives, in addition to the Chinese characters slipped under the door, a stream of numerals that appear, say, on a ticker tape in a corner of the room. The instruction books are augmented to use the numbers from the tape as input, along with the Chinese characters. Unbeknownst to the man in the room, the numbers in the tape are the digitized output of a video camera (and possibly other sensors)."[9]Searle argues that additional information will not allow for the man to gain understanding of the Chinese characters.

The fourth typical response to the Chinese Room thought experiment is often described as the Brain simulator reply. In this response, we imagine that the computer operates in a different way than the original thought experiment computer does. Instead the program simulates the actual sequence of nerve firings that occur in the brain of a native Chinese language speaker. Because the computer works in the same way as the brain of a native Chinese speaker, it will understand Chinese.[10]

Searle argues that this changes nothing. He poses a variation on the brain simulator scenario, " suppose that in the room the man has a huge set of valves and water pipes, in the same arrangement as the neurons in a native Chinese speaker's brain. The program now tells the man which valves to open in response to input."[11]This would suggest that the man does not

understand how the pipes work and likewise the computer would not understand Chinese.

The final response to the Chinese thought experiment is commonly referred to as the other minds response. " How do you know that other people understand Chinese or anything else? Only by their behavior. Now the computer can pass the behavioral tests as well as they can (in principle), so if you are going to attribute cognition to other people you must in principle also attribute it to computers. "[12]Searle's responds that we assume that other people have minds when we interact with them; " just as in physics we assume the existence of objects." [13]

While there are many responses to the Chinese room thought experiment, there is a fundamental flaw with the conclusion made. If the computer does not understand something if it follows a set of instructions, then we as people do not understand many things that we would claim that we do. For example, take a mathematic formula such as the Pythagorean Theorem. Many people can take the formula and use it to find an answer. The people that use the formula use a set of instructions that they are told in school. However, according Searle's logic, following a set of instructions does not constitute understanding. Searle would claim that you need to know all parts (more than the instructions) of something to understand something. However, this is problematic. People do not know everything and therefore there is always something that we do not know and understanding cannot be complete. In addition even if we think we know everything about an object, there always is a possibility of something else that we do not know changing our understanding of the object. Therefore, it is impossible, according to the <https://assignbuster.com/chinese-room-thought-experiment-philosophy-essay/>



Chinese Room thought experiment, to know or understand anything with and confidence.

In conclusion, Searle's thought experiment analyzes Artificial Intelligence and understanding. Searle's thought experiment draws similarities between Artificial Intelligence and a man following a set of instructions to make appear as if he understands Chinese. However, Searle's claim that it is impossible to have what he refers to as strong Artificial Intelligence was met with much controversy. Since its creation in 1980, there are 5 main responses to counter Searle's claim. However, there are more problems with Searle's argument than those five responses. If we take his claim and continue it to its natural conclusion we have to conclude that it is impossible to understand anything. This conclusion is problematic because our intuition tells us that while we do not understand everything, we do understand some things. Never the less, Searle's thought experiment is a famous thought experiment that is very thought provoking and deserves its place as one of the most famous thought experiments.