Reflections on Searle's Chinese Room Argument

Samir Roy*

Abstract

The Strong AI Thesis argues that a computing machine, while simulating some intelligent behaviour with the help of an algorithm running on it, actually shows intelligent behaviour. According to this thesis, there is no essential distinction between a natural phenomenon and a simulation of that phenomenon. John Searle, an American philosopher, strongly opposed this view. He put forward an intuitively appealing argument known as the Chinese Room argument. In this article, the author reflects upon Searle's Chinese Room argument and suggests that a deeper understanding of the nature of the human consciousness is required to settle the dispute.

Keywords: Consciousness, Artificial Intelligence (AI), Simulation

1. Background

Men's endeavour to mechanize the thought process has a long history. The most conspicuous outcome of this effort is the modern digital computer. The versatile power of a computer inspired the scientists to develop complex systems to carry out tasks that were hitherto completely in the realm of human intellect. Thus developed the branch of computer science named Artificial Intelligence (AI). A few milestones in the early ages of AI are noted below.

In 1948 the first chess playing computer was built at MIT. In 1950 W.Garey Walter invented a device called a *tortoise.* The *tortoise* was capable of moving around the floor of a room automatically until its power went low. Then it could search for the power socket and once found, plug itself into the socket. It remained there until its battery charge was thoroughly replenished, then detached itself from the socket and continued its adventure around the floor. The most spectacular success came in 1965 with Joseph Weizenbaum's program ELIZA. Its full name was 'ELIZA - A Computer Program for Natural Language Communication Between Man and Machine'.

Actually, ELIZA simulated the responses of a nondirective psychotherapist, in the form of typed output. Here is a typical interaction.

PATIENT : I feel depressed.
ELIZA : Tell me why you feel depressed.
PATIENT : I think I cannot succeed at anything.
ELIZA : You think you can't succeed at anything.
PATIENT : Yes that's right.
ELIZA : Please tell me more about why you feel depressed.

At about the same time K.M. Colby developed a similar program to simulate a psychotherapist. This program was so successful that some patients preferred the computer to a human therapist. A sample conversation with Colby's program is as follows :

THERAPIST : How can I help you?
PATIENT : I don't know.
THERAPIST : Try to tell me something about it.
PATIENT : No.
THERAPIST : Can you elaborate?
PATIENT : I am sick and tired today.
THERAPIST : Do you have any idea why?
PATIENT : Yes, I am old.
THERAPIST : Please go on.
PATIENT : I seem to have lost interest in everything.
THERAPIST : Tell me more about such feelings.
PATIENT : I have some idea why I feel this way too.

* *Lecturer, Department of Computer Science & Technology*

THERAPIST : You have.

PATIENT : It seems that I miss having a family.

THERAPIST : What makes you feel you miss having a family?

PATIENT : There is a challenge about it.

THERAPIST : Why is that important to you?

This may create an uncanny feeling that a conversation between two human beings are really going on, but the technique followed by ELIZA or Colby's psychotherapist is surprisingly simple. Here is the brief outline.

1. Read the sentence uttered by the patient.
2. Search for some keyword that can be used as a cue.
3. Change the verb from first person to second person.
4. Add a tag like 'Please tell me more about . . .' if necessary.
5. In case the above technique fails, play back one of the canned sentences, such as 'What else do you want to tell me about this?' or 'Please go on' etc.

Such and other successful AI programs made the AI scientists extremely enthusiatic about the potential of AI. This culminated in the Strong AI thesis.

2. The Strong AI Thesis

The phrase *Artificial Intelligence* was coined by John McCurthy, a pioneer in the field of AI, in 1956 at a conference at the Dartmouth College. The conference was funded by the Rockfeller Foundations. The purpose as stated in the application for the Rockfeller Foundation Grant was as follows.

"The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be built to simulate it."

Two things are to be noted here. First, we are assuming that *intelligence* is algorithmic (....can in principal be so precisely described....). And then, we are merely *simulating* intelligence, not creating it.

This still holds good as the underlying philosophy of Artificial Intelligence. However, early success in devising systems that display intelligent behaviour made some AI scientist take a rather extreme point of view in this regard. They argued that not only these systems are actually intelligent and have minds, but a kind of mental qualities can be attributed to the logical functioning of any computational device. This point of view is known as the Strong AI thesis. Some early proponents of this kind of view are Minsky (1968), Fodor (1983), and Morovec (1989).

3. Searle's Chinese Room Argument

With this background John Searle, an American philosopher and logician, put forward an argument in the form of a thought experiment. A thought experiment is a kind of experiment that can only be carried out at a logical level but not in real, physical world. The fact that it cannot be performed in the physical world stems from human limitation, rather than a contradiction with the established laws of nature. Therefore, in order to have an insight into some natural phenomenon, scientists-philosophers often find it worthwhile to ponder over such thought experiments knowing well that they are not feasible. Einstein himself was a guru of formulating wonderful thought experiments. Now, let us consider Searle's argument.

Consider a person P locked up in a room, alone. The person P does not know a word of Chinese. The only language he knows is English. However, he has at his disposal a set of instructions, written in English, corresponding to every conceivable question written in Chinese, which tells how to write the answer to that question in Chinese only. From outside, the only way to communicate with the person inside the room is to write a question on a piece of paper and pass it inside the room through a small hole. The person inside the room communicates the answer by writing it on a piece of paper and passing it through the small hole to the

person outside the room. Now whenever a person wants to ask the person P a question in Chinese, all he has to do is to write the question (in Chinese) on paper and pass it on to P through the said hole and wait for the answer. P, the person inside the room, receives the question. He cannot make head or tail of it, because he is absolutely ignorant of the Chinese language. But that does not matter. He carefully reads the instruction (written in English) regarding what kind of pictures (i.e Chinese alphabet) has to be laid down on paper corresponding to the patterns (i.e the question in Chinese he has obtained from outside) and carries out the instruction faithfully. Then he passes on the paper on which he has written the answer in Chinese (without understanding a word of it) to the person outside.

Therefore, in effect, from outside whatever question is being asked to the person P in Chinese, is being answered in Chinese only. Apparently, there is no way to distinguish the behaviour of the person inside the room from a person who really knows Chinese. Now, Searle asks, should we say that the person P *knows* Chinese? From the point of view of Strong AI thesis, the person P *knows* Chinese. However, our intuition says that there is something missing in the person P that makes us reluctant to certify his Chinese knowledge. What is that something which is missing in the person P?

4. A Mystic Reality

However reluctant we may be to admit, the truth is if we consider the person P to be a black box and stick to judge on the basis of his behavioural manifestations only, that something which is missing is nothing but nothing. Problem arises when we try to establish a parallelism between our own experience of knowing something with the inner state of mind of the person P with regard to his knowledge of Chinese. As soon as we do that, we go beyond the outword manifestation of the behaviour of a person and try to peep into his inner self. In other words, instead of looking from outside, we look from inside, we judge his personal *experience* rather than his behaviour. But what is *experience,* after all? And what is that mysterious entity who experiences? Before we ponder over this mystic reality, let us try to make the point more explicit with the help of another experiment.

The author's intention is to make the reader doubt whether the description of a chain of physical events is sufficient to capture the essence of human experience. Suppose I show you a patch of blue on a white piece of paper and ask you what colour is this. You answer by saying 'blue', provided you are not colour blind. Now let us trace the chain of events that occurred in this experiment.

Some light ray fell on the blue patch, most of which was absorbed by it. Only parts, a particular frequency corresponding to blueness, are reflected and that electromagnetic wave with a certain frequency entered your eyes. There are innumerable cells on your retina. Some of them were excited in a particular way and that excitation, in the form of electrical charging and discharging, was brought to a certain region in your brain through a chain of nervous cells. Simultaneously, some sound waves vibrated your ear drums and that vibration was also brought to your brain. Some complex electrical phenomena occurred inside some billions of interconnected neurons inside your brain which ultimately resulted in some signals to be sent to your vocal chord to vibrate in a certain way so that the word 'blue' is pronounced.

Now in this complex chain of events, where is the place of the 'blueness' that you have experienced? If we make a perfect robot which, when presented with the same patch of blue, will undergo similar chain of events inside some extremely sophisticated electronic circuits that result in the utterance of the word 'blue' through a perfect speaker, would we say that the robot has *seen* the blue patch? Or to be more precise, has it undergone the *experience* of seeing the blue patch?

It is not yet certain whether the essence of human experience can be captured within the positivistic methodological frame-work of science. Roger Penrose

opined that a new physics of consciousness is required for that. However, the question is older than the modern Galilean science. Through the ages, man tried to know the *knower* within himself. The fallacy is, consciousness is naturally inclined to focus its attention on an outer reality. To explore consciousness itself, one has to cleanse the consciousness of the outward experiences and bring it back to its purest form. This is tantamount to experiencing nothingness, i.e, a state when one is conscious (in contrast to sleep) but conscious of nothing. Experiencing nothingness is the necessary prerequisite for a proper understing of the subject-object relationship. It seems obscured, but there are specific ways to achieve this. But that is a different story.

5. Conclusion

This article presents a kind of argument, known as the *Strong AI* thesis, that claims that there is essentially no distinction between a mental phenomenon and its simulation on a computing device. Though from a functionalist point of view, the Strong AI thesis is almost indisputable but our intuition does not subscribe to this thesis. Philosopher John Searle (1987) contended the Strong AI Thesis with the help of his famous Chinese Room argument that has a strong intuitive appeal. Reflections on Searle's Chinese Room argument takes us back to the age-old problem of the relation between the subject and the object, the nature of human consciousness and deep mystries involving that reality.

The scientist is not a person who gives the right answers, he is one who asks the right questions.

— Claude Lavi Strauss