'CAN A MACHINE THINK?'

Saikat Bandyapadhyay
Dept. of Philosophy, Hiralal Majumdar Memorial College For Women
Dakshineswar, Kolkata, West Bengal

Abstract

The question 'Can machine think?' has been addressed in 1950 by Alan Turing with a proposed test, the Turing Test (TT), where a computer is to answer questions asked by humans. If the answers from the computer are not distinguishable from the answer coming from humans, the computer passes the Turing Test. In 1980 the validity of the TT has been challenged by John Searle in his paper 'Minds, Brains and Programs' published in the 'The Behavioral and Brain Science'. In the paper Searle sets out the Chinese Room Argument (CRA), which is one kind of objection to the Turing Test. It is also an attack on Strong Artificial Intelligence (Strong AI) Theory. In this paper first I would like to discuss about the Turing Test. After that views of Strong AI are to be examined and in the third section I shall focus on The Chinese Room Argument raised by Searle.

Keywords: *Machine, Think, The Turing Test, Chinese Room Argument, Artificial Intelligence*.

Introduction

The phrase 'The Turing Test' is most properly used to refer to a proposal made by Alan Turing (1950) as a way of dealing with the question whether machine can think. According to Turing, the question whether machine can think is itself 'too meaningless' to deserve discussion. However, if we consider the more precise question whether a digital computer can do well in

certain kind of game that Turing describes as 'The Imitation Game', at least then in Turing's view we do have a question that admits of precise discussion.

Turing in his paper 'Computing Machinery and Intelligence' sketches the following kind of game. Suppose that we have a person, a machine and an interrogator. The interrogator is locked in a room separated from the person and the machine. The interrogator is like a judge can converse with both the person and the machine by typing into a terminal. Both the person and the machine try to convince the interrogator that they are the human. On the other hand interrogator's task is to find out which of the two contestant candidates will be the machine, and which will be the human, only by their answer to questions. If the interrogator cannot make a decision within a reasonable time, then the machine is considered to intelligent and wins the game.

The basic idea of the game is that the interrogator attempts to determine the sex of one contestant by asking questions and receiving answers in writing. The goal of at least one contestant answering these questions is to cause the interrogator to make the wrong determination. No information is available to an interrogator other than the written answers, and at least one of the contestants answering questions is not obliged to tell the truth.

The Turing test, in its original form, is to replace by a machine one of the contestants of the imitation game who is not required to be truthful. If the results of the game remain unaffected by the presence of this machine, then this machine is said to be capable of thought. In other words, a machine that is indistinguishable from a human being solely on the basis of 'written' interaction is considered to be capable of thought. About this game Turing says, " I believe that in fifty years time it will be possible to program computers, with a storage capacity of about 10 , to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning . . . I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machine thinking without expecting to be contradicted."

It is important to note here that Turing paper is pretty informal, and, in some ways rather idiosyncratic. In his paper he himself raises many possible objections against his theory and tries

to confute them. He mentions those objections as follows: 1) The Theological Objection, 2) the 'Heads in the Sand' objection, 3) The Mathematical Objection, 4) The Argument from Consciousness, 5) Arguments from Various Disabilities, 6) Lady Lovelace's objection, 7) Argument from continuity in the Nervous System, 8) The Argument from Informality of Behavior and 9) The Argument from Extrasensory Perception.

There are many different objections to the Turing Test which have surfaced in the philosophy of mind during the past sixty years, nut all of these we cannot discuss. However, there is one argument—The Chinese Room Argument- that is mentioned so often in connection with the TT. The Chinese Room Argument is being considered as an attempt for refuting the TT.

Turing's article on the Imitation Game provides an elegant operational definition of Intelligence. The essence of the test proposed by Turing was that the ability to perfect stimulate unrestricted human conversation would constitute a sufficient criterion for intelligence. This way of defining intelligence, for better or for worse, was largely adopted as of the mid-1950's implicitly if not explicitly, as overarching goal of the nascent field of artificial intelligence.

The Strong Artificial Intelligence theory

In 1980, Searle publishes the paper 'Minds, Brains and Programs' in which he tries to refute the views of Strong AI theory along with Turing's conclusion. In his paper he makes a distinction between 'strong' AI and 'weak' AI. According to him strong AI is the view that the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be said to understand and have other cognitive states. That is, computers do understand, have cognitive states and can think. By contrast, 'weak' AI is the view that computers are merely useful in psychology, linguistic and other areas, in part because they can stimulate mental abilities, weak Ai makes no claim that computers actually understand or are intelligent. In this paper Searle has no objection to the claim of weak Ai. But he has a serious grievance against the theory of strong AI who claims that the appropriately programmed computer literally has cognitive states and that the programs thereby explain human cognition.

To discard the observation of strong AI, Searle considers a programs SAM (Script Applier Mechanism) made by Roger Schank in 1977. Searle describes the Shank's program as follows—

"*. . . the aim of the program is to stimulate the human ability to understand stories. It is characteristic of human beings' story-understanding capacity that they can answer questions about the story even though the information that they give was never explicitly stated in the story. Thus, for example, suppose you are given the following story: 'A man went into a restaurant and ordered a hamburger. When the hamburger arrived it was burnt to a crisp, and the man stormed out of the restaurant angrily, without paying for the hamburger or leaving a tip.' Now, if you are asked ' Did the man eat the hamburger?' you will presumably answer, ' No, he did not.' Similarly, if you are given the following story: ' A man went into a restaurant and ordered a hamburger; when the hamburger came he has very pleased with it; and as he left the restaurant he gave the waitress a large tip before paying his bill,' and you are asked the question, 'Did the man eat the hamburger?', you will presumably answer,' Yes, he ate the hamburger'. Now Schank's machines can similarly answer questions about restaurants in this fashion. To do this, they have a 'representation' of the sort of information that human beings have about restaurants, which enables them to answer such questions as those above, given these sorts of stories.' When the machine is given the story and then asked the question, the machine will print out answers of the sort that we would expect human beings to give if told similar stories. Partisans of strong AI claim that in this question and answer sequence the machine is not only stimulating a human ability but also*

1. *That the machine can literally be said to understand the story and provide the answers to questions, and*
2. *That what the machine and its program do explain the human ability to understand the story and answer questions about it."*

These are exactly the claims that Searle likes to refute with his Chinese Room Argument. The same argument would apply to Winogrades SHARDU, Weizenbaum's ELIZA, and indeed any Turing machine stimulation of human mental phenomena.

The Chinese Room Argument

The argument is basically a thought experiment presented by John Searle. Suppose Searle is a monolingual English speaker, cannot understand any Chinese and he is locked in a room which has a input-output window. Now, he is given 'a large batch of Chinese writings' plus 'a second batch of Chinese scripts' and ' a set of rules' in English for correlating the second batch with the first. The rules correlate one set of formal symbols with another set of formal symbols; here 'formal' means he can identify the symbols entirely by shapes. A third batch of Chinese symbols and more instruction in English enable him to correlate elements of this third batch with elements of the first two batches. Actually first batch is 'a script', the second batch is 'story' and the third batch is 'questions' and the rule in English is 'program' but Searle in the room does know nothing of it.

Now, after coming the script, story and questions, following the rule he manipulates the characters and produce a reply in terms of Chinese symbol which he pushes through the output window.

Suppose further, the Chinese answer that Searle produces are very good. In fact, so good, no one can tell that he is not a native Chinese speaker. Producing answers ' by manipulating uninterpreted formal symbols it seems as far as the Chinese is concerned, he simply behaves like a computer, specifically like a computer running Schank and Abelson's(1977) 'Scripts Applier Mechanism' story understanding computer.

But in imagining himself to be the person in the room, Searle thinks it is quite obvious that he does not understand a word of Chinese stories, actually he understands nothing except doing same formal manipulation. So Searle concludes, 'Schank's computer has nothing more that I have in the case, where I understand nothing.'

Searle argues that computers can never understand because computer programs are purely syntactical with no semantic. So what the strong AI is claiming false. Understanding is actually a biological phenomenon and only something with the same causal powers as brains can have.

Conclusion

We have already seen that Searle argues against the strong AI by stating the Chinese Room Argument and concludes that Computationalism is false. To him, strong AI can only be achieved by a computer having the same causal relations as the brain. After Searle, S. Harnad, in his paper ' What's wrong and Right About Searle's Chinese Room Argument?" summarizes the CRA, agrees with it by refuting Computationalism, and counters a few comments on the CRA and that 'there are still plenty of degrees of freedom in both hybrid and non-computational approaches'. Another important point is P.M. Churchland and P.S. Churchland in their paper' Could a Machine Think?' argue the Chinese room argument is false and so it is not proven that mentality cannot be achieved by pure symbol manipulation. However due to performance failures of classical AI and specific characteristics of brains, they think that "classical AI is unlikely to yield consciousness machines, but that systems that mimic the brain might."

In fine, this paper is about the Turing Test, strong AI and the Chinese Room Argument. For each of these terrains, there are lots of problems, loads of papers and articles but no consensus. This paper has only given an overview on some of these. More research, especially in the domain of cognitive science and philosophy of mind will helpfully bring some light on these issues.

References

1.Boden, Margaret A (1988), Escaping from the Chinese Room In the Philosophy of Artificial Intelligence, ed. Margaret Boden, 89-104, New York: Oxford University Press.

2.Boden, Margaret A (1988), Computer Models of the mind, Cambridge: Cambridge University Press.

3.Boden, Margaret A.(2006) Mind as Machine: A History of Cognitive Science, Oxford University Press.

4.Churchland, P.M. & Churchland, P.S., 'Could a Machine Think?' Scientific American 262.1: 32-37. 1990.

5.Fisher, John A. (1988), The Wrong Stuff: Chinese rooms and the nature of understanding, Philosophical Investigations 11(4): 279-299.

6.Harnad, S, (2001) 'What's Wrong and Right About Searle's Chinese Room Argument? In: M. Bishop & J. Preston (eds.) Essays on Searle's Chinese Room Argument. Oxford University Press..

7.Heil, John (1985), Philosophy of Mind: A Contemporary Introduction, New York: Routledge publication.

8.Putnam, Hillary(1975) The meaning of 'Meaning', Mind, Language and Reality, Cambridge: Cambridge University Press.

9.Schank, Roger C. (1977) Natural Language, Philosophy and artificial Intelligence, IN Philosophical Perspective in Artificial Intelligence, p.196-224, Brighton, Sussex: Harvester Press.

10.Searle, John, R.(1980) 'Minds, brains and programs', Behavioral and Brain Science, p. 417-457. Stanford Encyclopedia of Philosophy – The Turing Test-http://plato.stanford.edu/entries/turing-test

11.Turing, M.A., (1950) 'Computing Machinery and Intelligence', Mind 49: pp. 433-460