Turing Machines and The Chinese Room

T L Hurst
1st Dec 2012

Abstract

This paper looks at the Turing test, Turing machines and Searle's 1980 thought experiment "The Chinese Room". It concludes that the Chinese Room scenario does not support his assertion that a modern computer is neither capable of understanding, nor the "right stuff" to support intent.

Introduction

Searle's 1980 paper "Minds, Brains and Programs"[1] has been the subject of much discussion by researchers into AI, as it suggested that the expectations of strong AI at the time were highly over-optimistic. We take a look at Searle's argument and some prominent replies, but first we need to discuss the Turing test and Turing machines, as these are fundamental to the argument...

The Turing Test

The Turing test[2] is a test of a machine's ability to exhibit intelligent behaviour. There are a number of variants, but the idea is for two hidden subjects to converse via a computer keyboard and screen for a fixed period of time. A judge can see the comments made by both, and has to decide which is human. If the judge cannot reliably distinguish a machine from a human being, the machine is said to have passed the test. However, I would suggest that there is, potentially, at least two ways of succeeding in the test:

a. By strategies designed to conceal the fact that the computer does not actually understand the input from the other subject (i.e. by subterfuge on the part of the programmers).

b. By enabling the computer to understand the semantic content of the conversation, so that it actually makes semantically appropriate responses.

Note: So far, it seems that computers which have "passed" the Turing test have done so by using subterfuge. However it does not necessarily follow that computers cannot understand semantic content. Indeed it is arguable that computers do already have a limited form of "understanding", in the sense that they can, and do, respond semantically correctly to their machine language. That functionality is hard coded in the CPU (central processing unit). Hence it may be argued that a machine without a processing unit of some sort is necessarily incapable of passing the Turing test.

Turing Machines

Turing machines[3] are abstract descriptions of the most basic elements of a computational device. Practical Turing machines may be constructed from sundry items such as stones and toilet rolls, and although, in principle, a Turing machine can perform any task that a digital computer can, we need to bear in mind that not all Turing machines are equal:

- Turing O machines require input from an operator (colourfully known as an "oracle"). The abacus and slide rule are examples of O machines. An O machine, by itself, is incapable of passing the Turing test. Indeed it is incapable of any action without the operator.

- A (automatic) machines are able to complete an automated sequence of actions independent of human operators. An A machine that is capable of emulating any other Turing machine is known as a UTM (universal Turing machine). Modern digital computers are UTMs.

Searle's Objective

Searle's stated objective was:

> *"My discussion here will be directed at... the claim that the appropriately programmed computer literally has cognitive states and that the programs thereby explain human cognition... I will consider the work of Roger Schank and his colleagues at Yale (Schank and Abelson 1977)... But nothing that follows depends upon the details of Schank's programs. The same arguments would apply to Winograd's SHRDLU (Winograd 1973), Weizenbaum's ELIZA (Weizenbaum 1965), and indeed any Turing machine simulation of human mental phenomena."*

Schank's Progam

Searle describes Schank's program as follows:

> *"The aim of the program is to simulate 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 burned 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..."*
>
> *"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.

The Chinese Room

Searle introduces the Chinese room thought experiment, in which he envisioned being locked in a room and being passed a story and a set of questions about the story, both of which were in Chinese (which he did not understand). He was also passed a series of instructions in English (which he did understand), together with a database of information. Using the instructions and the database, he could identify Chinese characters in the questions, and produce semantically appropriate replies to the questions that would convince a Chinese speaking person that they were conversing with a person who understood Chinese.

Searle also introduces a set of stories and questions in English, which he replies to in English, to contrast with the way that he produced the Chinese replies. He then compares the instructions to a computer program, and himself to a computer executing the program, and says:

> *Now the claims made by strong AI are that the programmed computer understands the stories and that the program in some sense explains human understanding. But we are now in a position to examine these claims in light of our thought experiment.*
>
> *As regards the first claim, it seems to me quite obvious in the example that **I do not understand a word of the Chinese stories**. I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, and I can have any formal program you like, **but I still understand nothing**.*
>
> *As regards the second claim, that the program explains human understanding, we can see that the computer and its program do not provide sufficient conditions of understanding since the computer and the program are functioning, and **there is no understanding**.*

Searle's Claims re Turing Machines

> *First, the distinction between program and realization has the consequence that the same program could have all sorts of crazy realizations that had no form of intentionality. Weizenbaum (1976, Ch. 2), for example, shows in detail how to construct a computer using a roll of toilet paper and a pile of small stones.*
>
> *Similarly, the Chinese story understanding program can be programmed into a sequence of water pipes, a set of wind machines, or a monolingual English speaker, none of which thereby acquires an understanding of Chinese.*
>
> *Stones, toilet paper, wind, and water pipes are the **wrong kind of stuff to have intentionality** in the first place-only something that has the **same causal powers***

> ***as brains can have intentionality****-and though the English speaker has the right kind of stuff for intentionality you can easily see that he doesn't* ***get any extra intentionality by memorizing the program, since memorizing it won't teach him Chinese****.*

Searle then proceeds to discuss some prominent counter arguments. The most important of these are:

1. The System Reply

> *"While it is true that the individual person who is locked in the room does not understand the story, the fact is that he is merely part of a whole system, and the system does understand the story. The person has a large ledger in front of him in which are written the rules, he has a lot of scratch paper and pencils for doing calculations, he has 'data banks' of sets of Chinese symbols. Now, understanding is not being ascribed to the mere individual; rather it is being ascribed to this whole system of which he is a part."*

Searle's reply is:

> *"Let the individual internalize all of these elements of the system. He memorizes the rules in the ledger and the data banks of Chinese symbols, and he does all the calculations in his head. The individual then incorporates the entire system. There isn't anything at all to the system that he does not encompass. We can even get rid of the room and suppose he works outdoors. All the same, he understands nothing of the Chinese, and a fortiori neither does the system, because there isn't anything in the system that isn't in him. If he doesn't understand, then there is no way the system could understand because the system is just a part of him."*

2. The Robot Reply

> *"Suppose we wrote a different kind of program from Schank's program. Suppose we put a computer inside a robot, and this computer would not just take in formal symbols as input and give out formal symbols as output, but rather would actually operate the robot in such a way that the robot does something very much like perceiving, walking, moving about, hammering nails, eating, drinking—anything you like... Such a robot would, unlike Schank's computer, have genuine understanding and other mental states."*

Searle's reply relates back to the Chinese room scenario...

> *...notice that the same thought experiment applies to the robot case. Suppose that instead of the computer inside the robot, you put me inside the room and, as in the original Chinese case, you give me more Chinese symbols with more instructions in English for matching Chinese symbols to Chinese symbols and feeding back*

> *Chinese symbols to the outside... I don't understand anything except the rules for symbol manipulation. Now in this case I want to say that the robot has no intentional states at all; it is simply moving about as a result of its electrical wiring and its program. And furthermore, by instantiating the program I have no intentional states of the relevant type. All I do is follow formal instructions about manipulating formal symbols.*

3. The Brain Simulator Reply

The brain simulator proposes simulating the function of the brain at the synapse level. Searle's reply again relates back to the Chinese room scenario...

Searle's Closing Comments

In his closing comments Searle says:

> *"I see no reason in principle why we couldn't give a machine the capacity to understand English or Chinese, since in an important sense our bodies with our brains are precisely such machines."*

However he adds:

> *"But I do see very strong arguments for saying that we could not give such a thing to a machine... where the operation of the machine is defined as an instantiation of a computer program."*
>
> *"...the main point of the present argument is that no purely formal model will ever be sufficient by itself for intentionality because the formal properties are not by themselves constitutive of intentionality, and they have by themselves no causal powers except the power, when instantiated, to produce the next stage of the formalism when the machine is running."*
>
> *"...any other causal properties that particular realizations of the formal model have, are irrelevant to the formal model because we can always put the same formal model in a different realization where those causal properties are obviously absent. Even if, by some miracle, Chinese speakers exactly realize Schank's program, we can put the same program in English speakers, water pipes, or computers, none of which understand Chinese, the program notwithstanding."*

Critique

Searle stresses that the computer program does not enable the operator to understand Chinese, and hence he does not gain any extra intentionality. However the program was never intended to do that, as the room provides appropriate replies without it. We can gain an insight into how the

room achieves that if we put a monolingual Arabic speaker alone in the Chinese room. He would necessarily fail to produce appropriate replies, because he would neither understand the story etc. in Chinese, nor the instructions in English. Hence it seems that the room, which is a Turing O machine, requires an operator's semantic understanding of English or Chinese to succeed.

However, the scenario is predicated on the computer being able to produce such replies. Therefore it would seem appropriate to equate a modern digital computer to a Turing O machine, with its operator. If so, whilst accepting that "*stones, toilet paper, wind, and water pipes are the wrong kind of stuff to have intentionality*", it would appear that Searle has not shown that the same applies to a modern digital computer.

References

[1] J R Searle - Minds, Brains, and Programs, 1980, "http://cogprints.org/7150/".
[2] Wikipedia - Turing test, "http://en.wikipedia.org/wiki/Turing_test".
[3] Wikipedia - Turing machines, "http://en.wikipedia.org/wiki/Turing_machine".