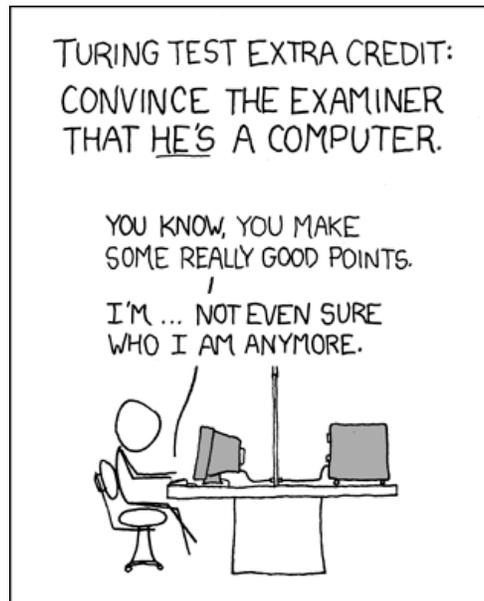


## Can Machines Think?: The Turing Test

### Language and Thought



**Turing Test:** Test proposed by Alan Turing (1950) as a way of addressing the question of whether machines can think. Turing proposed the following game: there is a person, a machine, and a (human) questioner. The questioner is in a room separate from the other person and the machine. The object of the game is for the questioner to determine which of its two respondents is a person, and which is a machine.

**Bold Proposal:** If a machine can pass the Turing Test, then it can think.

*Note that this proposal is vague, and can be fleshed out in different ways. How long does the game need to go on? How many examiners does the machine need to convince?*

### **Searle's Chinese Room Thought Experiment.**

*Background:*

**Intentionality** = the feature – possessed by e.g. words and mental states – of being “about” something (that is, representing something, referring to something).

Searle's argument is intended to show that merely implementing a (sufficiently advanced) computer program isn't sufficient for intentionality. Given the assumption that intentionality is required for thinking, it will – if successful – show

that merely implementing a sufficiently advanced computer program isn't suffice for thinking. In other words, Searle aims to show that passing the Turing Test isn't sufficient for thinking.

*The Chinese Room Scenario:*

A man who does not know how to speak or read Chinese – Searle – is in a room. There are two small windows in the room. Through window A, he receives slips of paper with messages written in Chinese (the input). Searle then consults a very big instruction manual, which specifies how he should respond to a given input. He then writes down the specified output, which is also in Chinese, on another slip paper, and passes it through window B.

Such a person could write down the output and – if his instructions were adequate – these would pass the Turing Test: it would seem like he's a competent speaker of Chinese. But, Searle claims, this person wouldn't count as understanding Chinese.

More generally, Searle takes the lesson to be that the mere ability to manipulate symbols – produce one set of symbols in response to another – is insufficient for understanding/intentionality. And, according to Searle, all computer programs do is manipulate symbols. They don't understand these symbols any more than the man in the Chinese room understands Chinese. Searle puts the point by saying that syntax is not sufficient for semantics.

*Some Replies:*

**Systems Reply.** The man doesn't understand Chinese, but he's part of a larger system – including the room – which does understand Chinese.

*Searle's response:* "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... 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."

**Robot Reply:** Searle's right that neither the man nor the whole system is conscious/thinking. But if we put a digital computer in a robot body, together with sensors (e.g. video cameras and microphones), the resulting machine would be able to think.

**Searle's response:** All the sensors do is provide the computer with additional information. But this is itself merely formal (syntactic) information, and hence won't be sufficient for intentionality/thinking. To see this, imagine that the man receives – in addition to the Chinese characters – a stream of binary digits from outside the room. Imagine that the man's

instruction book is enriched with instructions for manipulating these digits. Unbeknownst to the man, these digits are the digitized outputs of a video camera (and possibly other sensors). Still, this won't enable the man to associate meanings with the Chinese characters.

Note we can distinguish between two theses:

**Thesis 1:** There is a computer program such that any (possible) computer running this program counts as thinking.

**Thesis 2:** There is a computer program such that any (possible) computer running this program that's related to the world in the right ways (e.g. via sensory input) counts as thinking.

In responding to the Robot Reply, it seems Searle is trying to combat both theses.

**Luminous Room Analogy:** Patricia and Paul Churchland offer the "Luminous Room" as a parody of the Chinese Room argument. They ask us to imagine someone – call them Earle - who believes magnetism and electricity are forces. Earle might then try to argue that no electromagnetic forces could ever suffice for light, as follows: "Imagine someone in a dark room waves a magnet around. No matter how fast they wave the magnet, the room will stay dark. Hence light can't be an electromagnetic force."

The objection here is that the claim, *Electromagnetic forces are not sufficient for light*, is an empirical claim, and a false one at that. It is not a claim that can be demonstrated by a thought experiment like the Luminous Room. Likewise, the objection runs, Searle's claim, *syntax is not sufficient for semantics*, is an empirical claim – one that cannot be demonstrated by a thought experiment such as the Chinese Room.

**Qs:** Do you find this a convincing objection? Do you find Searle's replies to the earlier objections persuasive?