

**Reasoning and Communicating**

Oct 10, 2004

***'Let us say that you are confronted with an intriguing device. What should it take to convince you that the device can think?'***

This essay will attempt to address the issues presented by the question above - that is; what observable criteria must be fulfilled in order to show beyond reasonable doubt that thought is occurring in a device of any nature? We will assume that the device is some form of digital computer, given firstly the fact that computers can be programmed to mimic almost any set of circumstances, and secondly the readily available nature of existing research into computers and their relation to thought. In attempting to answer the question, we must decide what the functions of thought are, and how to observe them, while also recognising the inherent flaws in observable characteristics. We must hypothesise about the nature of the computations performed by our device. We will consider some established tests of intelligence with regard to computers, and the appropriate counter-arguments for these tests, along with a critical analysis of parts of these counter-arguments. Finally, there will be a short reflection with regard to some alternative forms the device could take.

Let us first examine the definition proposed in the introduction. It is important not to get sidetracked by the subjective nature of the device in question; as Turing states, when considering an engineered human child:

"We would not be inclined to regard it as a case of 'constructing a thinking machine'"<sup>1</sup>

Any conclusions based on a device of finite capabilities will only be true of that particular device. Hence, when considering 'a device', we must take it to mean 'a man made device of any nature'.

By observing how the device interacts with its environment, we can guess whether a logical process is going on behind the actions. Some other observable functions of thought are problem solving, survival instinct, and more importantly, rationalisation. If a device were to be able to show evidence of these factors, we would be closer to being able to assume that the device thinks. Indeed, this is the theory we can infer is behind Turing's ideas on thinking machines.<sup>2</sup>

In order to ascertain whether these functions are being performed, we must assume that the device is capable of communicating with us, and furthermore, that we are capable of understanding the device's communications. Turing suggests a series of conversations, during which questions are asked of the device and of a human, concerning various topics ranging from mathematics to problem solving, contextual analysis and so forth. If the questioner cannot tell which is the device and which is the human, Turing believes that the device can be considered to be capable of thought<sup>3</sup>. This approach is appealing, as it brings the possibility of an intelligent machine closer to realisation, indeed computer programs have been produced that are capable of passing the Turing test<sup>4</sup>. However, although it is possible to produce devices that these functions are observable in, we would not be correct in saying that we had before us a thinking device - the underlying force behind all the above criteria is *understanding*, a feature whose absence would negate any assumption that the device could think, and whose presence is not proved by simple imitation. Such programs as are based on Turing's theories cannot be said to understand the nature of the tasks they perform, they simply pass input through a series of algorithms in order to produce output. Understanding, if present at all, would be an attribute of the algorithms rather than a feature of the machinery.

---

1 'Computing, Machinery and Intelligence', 1964, p7.

2 'Computing, Machinery and Intelligence', 1964.

3 This technique is known as the 'imitation game', or 'Turing test'.

4 Eliza (1965), Online AOL Tech support (c. 2002), and others.

***'Let us say that you are confronted with an intriguing device. What should it take to convince you that the device can think?'***

We can look to Searle<sup>5</sup> for a compelling argument that such 'Turing passed' devices do not necessarily understand. Searle considers programs that can correctly answer contextual questions that have not been explicitly answered, and suggests that such programs do nothing more than 'formal symbol recognition'; they take an input of which they know not the nature, perform certain 'hard wired', or 'hard coded'<sup>6</sup> functions on that input, and produce output, also of an unknown nature to them. At none of these stages can the program be said to possess understanding. Searle poses an exemplary experiment<sup>7</sup> involving a man who does not understand Chinese, being locked in a room with a set of Chinese characters. The man is passed two further (different) sets of Chinese characters, A and B, having been given instructions on how to match the patterns on the two sets of characters and return the correct characters from his original set. Suppose that set 'A' was a story, set 'B' a question, and the characters he passes back, the correct answer. The man would not have understood that he had just answered questions on a story, and in a similar way, we cannot assume that any different process is occurring in our device. Searle goes on to hypothesise that it is the actual physical and chemical make-up of the brain that produces intentionality. Another of Searle's key arguments is that any simulation of understanding cannot equivocate true understanding, as he states:

"To confuse simulation with duplication is (a) mistake"<sup>8</sup>

We can be sure that the man in Searle's experiment does not understand what he is doing, however, it can be argued that the combination of the man, the rules, the characters, and the semantics of the symbols constitutes a system that understands what it is doing<sup>9</sup>, and that the man, having memorised the system or not, is always only a part of that system.

The implications of systems theory which concern this topic are twofold; firstly, our device by itself still cannot be said to understand, but instead comprises part of an overall understanding system. Secondly, there are a number of wider implications – if we accept this theory, we are bound to accept that other circumstances may also constitute systems of understanding, and vice-versa; if we accept that the systems theory applies in some cases, we cannot rule out the idea that it may also apply for electronic brains.

We must consider Searle's reply to the effect that the man in his experiment may memorise all the elements of the system and still not understand. There is an objection to this argument which must be raised, that while the man himself does not understand, the argument that the system understands is surely still valid; the man can still be said to be separate from the system that he has internalised by memory. There is no distinction between the man's understanding of English and the system's understanding of Chinese in a way that would exclude the 'English understanding' part of the man from 'formal symbol recognition'. In essence, if we accept that the Chinese system does not understand, simply because it merely processes formal symbols, then we must acknowledge the possibility that the part of the man that understands English also does nothing more than symbol recognition. With the distinction gone, we can theorise that both the English and Chinese 'subsystems' do in fact understand, while recognising that symbol recognition is a part of understanding, and that the man as a distinct entity will never understand.

---

5 'Minds, Brains and Programs', MIT Press, 1981.

6 A 'Hard Coded' function is one whose inner workings are static – the algorithms used cannot be changed.

7 The 'Chinese Room' experiment - Searle - Minds, Brains and Programs.

8 'Minds, Brains and Programs', MIT Press, 1981, p302

9 "The Systems Reply (Berkeley)" - Searle - Minds, Brains and Programs.

***'Let us say that you are confronted with an intriguing device. What should it take to convince you that the device can think?'***

With Turing's imitation game countered by Searle, it is conceivable that the question is answered by default; that no man made device can think, simply because there is no separation of mind and brain, and therefore nothing other than a brain can possess a mind, by which we can infer that nothing other than a brain can think. This hypothesis is dubious, however. The material of which the brain is made is not unique. At the most fundamental level, it is atoms and electrons, so Searle's hypothesis should surely run as follows; anything that is comprised of material other than atoms and electrons cannot think. Such a hypothesis clearly leaves room for thinking machines, and therefore does Searle no good as a counter argument to the idea of such machines. To say that it is not the actual materials, but somehow the way in which the physical and chemical components interact that makes the brain unique, seems to suggest that one type of amalgamated system, a brain, does understand, yet another type of system, the one that, among other things, performs symbolic manipulation, does not. This is an obvious contradiction, and we are forced to conclude either that the brain as a system does not understand, or that the Chinese symbolic processor as a system does understand. Whichever conclusion is accepted, room remains for thinking machines; if the brain does not understand, then we must accept that it is simply the matter of which it is comprised that makes it understand, in which case there is no reason that a device could not mimic the structure of the brain. Alternatively, if the symbolic processor understands, then there already numerous examples of 'understanding systems'.

An obvious objection to this idea is the fact that the brain is not merely an arrangement of atoms and electrons, rather, it is an arrangement of certain chemicals<sup>10</sup>, the conclusion then being that unless a device is constructed of these particular chemicals, it cannot think. However, unless a single chemical by and of itself understands, we are still accepting that a system understands, rather than a distinct part of a system, hence, we must still accept the possibility that other systems, namely the Chinese Room system, understand.

Crane suggests that if syntax does not equal semantics, then perhaps syntax may gain semantics via interaction with the world. However, as he states;

"Nothing can think simply by being a computer"<sup>11</sup>

The reason being that without interaction, no understanding can take place. This seems to present another notion, that learning is a function of understanding – mere interaction would get us no closer to having a thinking machine. If it were possible to create a machine that could interact with its environment, learn about the nature of its environment, and modify its behavior accordingly, we can argue that we would be justified in saying we had a thinking machine. We could certainly no longer say of the machine that it simply processes formal symbols – the fact that it intelligently modifies its behaviour may show evidence of understanding.

In conclusion, we have seen a great deal of confusion regarding the nature of thought, and its subprocesses, such as understanding and learning. In order to determine whether a machine thinks, we must have definitive undisputed criteria of what constitutes thought, we must accept that if it not possible to determine if our fellow human beings can think, then we cannot assume any more of our hypothetical device.

What would it take to convince me that a device could think? It would be tempting, if faced with a device such as described above, that had broken out of the 'hard-coded' nature of its

---

<sup>10</sup>Various neurotransmitters, and so on.

<sup>11</sup>The Mechanical Mind, 1995

## Reasoning and Communicating

Oct 10, 2004

***'Let us say that you are confronted with an intriguing device. What should it take to convince you that the device can think?'***

functions, to say that the device was capable of thought; it could be said to possess a form of programmed evolution. However such a conclusion is highly questionable. The simple truth is that currently, nothing can convince me that a device can think, as there does not exist a definitive definition of thought; to be convinced that a device was capable of some undefined property would be foolish.

***'Let us say that you are confronted with an intriguing device. What should it take to convince you that the device can think?'***

**Bibliography**

- 'Minds and Machines', Prentice-Hall, 1964. A.R. Anderson (ed)  
- *section: Computing, Machinery and Intelligence, A.M. Turing*
- 'The Mechanical Mind', Penguin, 1995. T. Crane, ch 3.
- 'Mind Design', MIT Press, 1981. J Haugeland (ed)  
- *section: Minds Brains and Programs, J.R. Searle*
- 'Descartes', Oxford University Press, 1987. T. Sorell