The Mechanical Mind

Lecture 3: Computers and Minds

1. In this last lecture we return to the two questions with which we began: Can computers think? Do we think by computing? We shall start with the first question: **Can computers think?**

2. It is very tempting to attribute thought to computers, given the way they *behave*. They behave in ways that can very well be explained by employing common sense psychology: "My computer hates me!". If such explanations work so well, surely that's good reason to think that computers think?

   Also, computers are like minds in that they are both causal mechanisms that involve representations.

3. But there are two famous arguments that computers *cannot* think, no matter how sophisticated they are.

4. Hubert Dreyfuss argues that thinking requires abilities that computers do not have, and by their very nature can *never* have. Computers have to follow rules (recall that we are taking computers to be Turing machines), but thinking can never be captured by a system of rules, no matter how complex – rules can never capture common-sense knowledge. (Consider the bus driver who sticks to the rules, even though a passenger is dying.)

5. We can think of Dreyfuss as claiming that a computer could never pass the *Turing test* (Turing, 1950). What is the Turing test? The test is whether or not a person who is having a conversation with a computer and a person could distinguish the computer from the person; if he cannot, then the computer passes the test.

   Passing the Turing test is plausibly a necessary condition for thinking, and Dreyfuss is claiming that a computer never could. (See 'The Turing Test' in SEP.)

6. John Searle argues that even if a computer *could* pass the Turing test, it would not be thinking, but merely *simulating* thought (so passing the Turing test is not *sufficient* for thinking). The reason is that computers only manipulate symbols, without any understanding of what those symbols *mean*, and thought requires *understanding*.

   He illustrates with his *Chinese room* thought experiment:

   > Imagine Searle in a room with an input window through which come sheets of paper with questions in Chinese. Searle has a translation manual written in English which he uses to send answers in Chinese back through an output window, in such a way that the Chinese room 'computer' passes the Turing test.

   The Chinese room passes the Turing test (Searle grants), but since Searle doesn't understand Chinese the Chinese room is not *thinking*.

   a. *Objection*: *Searle* doesn't understand Chinese, but *the room* does.
   b. *Reply*: Get Searle to memorize the translation manual, so that it is *Searle* who passes the Turing test, and it is *Searle* who does not understand.
   c. *Objection to reply*: Then it is not obvious that Searle does not understand.

   (See 'Chinese Room Argument' in SEP, and Searle's *Minds, Brains & Science*, ch. 2)

7. Now let's turn to the second question: **Do we think by computing?**

8. It might not be plausible that *all* of our thinking proceeds by computation: is there computation involved when I believe that I have a pain in my toe? But it is at least plausible that we *reason* by computing. If reasoning is a process by which we move from certain beliefs to certain other beliefs, if those beliefs are representational states (as we are supposing), and if we do so in an algorithmic manner, then that sounds a lot like computing.

9. Is there any reason to think that is what happens when we reason? If the *language of thought hypothesis* is correct, then perhaps there is.

10. The LOT hypothesis (Jerry Fodor). There is a language of thought, sometimes called *LOT*, sometimes called *Mentalese*. What does that mean?

    When you have a thought, say that today is Wednesday, a sentence is *tokened* in your head, which has the same meaning as the English language sentence 'Today is Wednesday'. It is not a sentence of English – it is a sentence of Mentalese – but it has the same meaning as this sentence of English. And it is not tokened in the way that we often token sentences of English – as marks on paper or as vibrations of air particles. Rather, it is tokened as a pattern of neuron firings in the brain.

    Being a language, Mentalese has *syntax* and *semantics* and is *compositional* – the meanings of complex sentences are composed in a systematic way from the meaning of their simpler constituents. When you have the thought that today is Wednesday, there is a pattern of neurons in your brain that tokens the sentence 'Today is Wednesday'; this pattern has three constituents, which token the three constituents of the sentence – 'Today', 'is', and 'Wednesday'; the meaning of the whole is built up from the meanings of the parts. Compare this with a non-compositional language: flags on ships.

11. Why should we think the LOT hypothesis is true? Fodor has advanced two arguments:

    a. Thought is compositional: anyone who can think that Katie loves Jamie can thereby think that Jamie loves Katie; we can have infinitely many novel thoughts. The LOT hypothesis is the *best explanation* of this. (This is an *inference to the best explanation* style of argument).

    b. It is part of our common sense psychological theories that much of our thought involves *reasoning* about what to do and what to think (rather than free association of ideas). When we reason we move *validly* from one thought to another. Whenever one thing follows validly from another thing, it is because of their *syntactic form* (think of *modus ponens*). Since some of our thoughts follow validly from other of our thoughts, our thoughts must have syntactic form.

12. If the LOT hypothesis is true, then it is plausible that when we reason we do so by manipulating meaningful symbols (patterns of neuron firings) according to certain logical rules – that is, by following an algorithm. So it is plausible that at least *some* of our thinking proceeds by computation.