Functionalism and the Chinese Room

Minds as Programs

Three Topics

- Motivating Functionalism
- The Chinese Room Example
- Extracting an Argument

Motivating Functionalism

- Born of failure, to wit the failure of the Mind-Brain Identity Theory
- Also instigated by a flight from dualism.

Mind-Brain Identity Theory

- In its purest form, the reductive MBI holds that every mental property is *identical* with some single physical property (Type-Type Identity Theory—TTIT).
  - Thus, the property of being in pain is identical with the property of being neural state N^{237}.
  - Or the property of being a belief that Vienna is the most beautiful city in Europe is identical with the property of being neural state N^{2459}.
- This theory evidently founders on the multiple realizability (MR) of the mental:
  - Possibly, something is in pain though it is not in neural state N^{237}.
    - Possibly, that is, there are aliens, androids, angels . . .all of whom can be in pain.
  - So, pain is not identical with N^{237}.

Rethinking MR

- Perhaps we are at the wrong level with the MBI:
  - The property *being a belief* is rather like the property of *being a poison*: every poison is some stuff or other, but poison can take various physical forms.
  - So, poison is multiply realizable.
  - For any one stuff to qualify as a poison, it must simply play a certain role:
    - It must be *such as to cause grievous harm or death to an organism*.
      - Of course, different first-order stuffs can play that role.
        - arsenic, strychnine, hemlock, cyanide. . .

Perhaps. . .

- Instead of merely refuting reductive materialism, MR actually points the way to a defensible form of materialism—or at least a defensible sort of non-reductive materialism:
  - This is, namely, *functionalism*.
    - Perhaps mental states are functional states, rather like the states of computers.
      - It does not matter in what sorts of machines programs are realised.
        - It only matters that they be executed, in accordance with some specifiable rules.

Machine Functionalism

- A soda dispenser:

	S^0	S^1	S^2
Input: 50¢	Go to S^1	No output; Input: 50¢ Go to S^2	Output one soda; Go to S^0
Input: $1	Go to S^2		Output one soda; Go to S^0
Input: $2	Go to S^2		Output one soda + $1 Go to S^0

Soda Dispensers are MR

- Anything that can be in a state capable of playing the roles defined by this machine table can be a soda dispenser.
  - Soda dispensers can be of plastic, metal, fiberglass, wood. . .
  - Indeed, soda dispensers can be made of flesh and bones: you yourself (whatever else you may be) can be a soda dispenser.

Minds are MR

- Perhaps, then, minds are simply programs running on suitably sophisticated hardware, such as the brain; but it needn't run on a brain. Any suitably sophisticated hardware will do.
- Our view is, then, fully compatible with materialism. But it is not a version of the MBI.
- One might say that functionalism is materialistically adequate, without striving to be reductive.
  - Anything can be a mind—as long as it can realise the relevant program which defines the transitions involved in mental states.
  - So, e.g. the state of being in pain:
    - Pain is that state which: (i) tends to be caused by bodily injury; (ii) to produce the belief that something is unpleasant and so to be avoided; (iii) to cause a desire to engage in avoidance behaviour; (iv) to override other conscious mental states; and (v) to cause wincing, moaning, and shrieking.
    - Anything which can be in a state which plays that role qualifies as being in pain.
      - It matters not a bit what the system is made of—as long as the stuff in question can realise that condition.

Could a computer think?

- If functionalism is true, then it seems so.
- To have a mind is neither more nor less than to implement a certain computer program.
- Thinking *is* computation.
- That is, a suitably programmed computer could (would, in fact) do more than mimic human thinking.
  - Such a computer could actually think—that is, it could (would, in fact) have beliefs, hopes, fears, desires. . .

Passing the Turing Test

- Alan Turing proposed a simple answer to the question of whether a computer could think.
- Or, rather, having eschewed that question as ill-defined, proposed a test as to whether a computer could pass the imitation game:
  - First: "I propose to consider the question 'Can machines think?'" (RR, 391)
  - Second: "Are there imaginable digital computers which would do well in the imitation game?" (*RR*, 394)
  - If so, then they will have passed what is now known as the "The Turing Test":
    - If a computer can pass for human in unrestricted conversation, then we should grant that the computer is an intelligent being.
      - After all, this is our primary evidence for believing that *humans* are intelligent.

Turing's Presentation

- I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, 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 machines thinking without expecting to be contradicted. —Turing (1950)

The Chinese Room

- Searle argues that passing the Touring Test is *not* sufficient for being intelligent:
- 'Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese.' (Searle, 1999)

Weak and Strong Artificial Intelligence

- Weak AI: a suitably programmed computer models or simulates human understanding.
- Strong AI: a suitably programmed computer understands, as humans understand, and is conscious, as humans are conscious.

The CR Argument

1. If strong AI is true, then a computer running a program understands the semantic dimensions of that program.
2. If a computer running a program understands the semantic dimensions of a program, then the compiler in the Chinese room exhibits an understanding of Chinese.
3. The compiler in the Chinese room does not exhibit an understanding of Chinese.
4. So, strong AI is false.