

## Minds, Programs and Robots

### I. RECAP

Identity Theory: Mental states/events are brain states/events.

Behaviorism: Mental states/events are dispositions to behave in certain ways.

Eliminativism: There are no mental states/events.

Functionalism: Mental states/events are functional states of a complex system, i.e., they are defined by their causal role in a holistic psychology.

Identity theory leads to an implausible speciesism -- only us and creatures like us can really think and feel. To avoid this result some philosophers have proposed mental states should be identified not with physical states of the brain but something higher and more abstract, something that could be duplicated in all *kinds* of physical structure.

Pain, for example, would be defined by one of the columns in the table; it would be the state which is caused by such and such types of bodily injury and leads to such and such psychological and behavioral reactions. Any state that played exactly this functional role would be pain. Likewise for wanting a latte, experiencing a red afterimage, and so on.

### II. Problems with functionalism?

Question: Can machines think?

Why does this matter?

Ethical considerations: if so, perhaps they have rights.

What would it take for a machine to think?

What distinguishes a thinking machine (Data?) from an unthinking machine (pocket calculator, thermostat)?

Turing Test:

An "interrogator" uses text messaging to question a pair of "contestants" in another room: a human and a computer program. The interrogator's job is to distinguish the human from the computer program using only their answers. She may ask whatever questions she wants. The contestants' job is to convince the interrogator that they are human. If the interrogator cannot distinguish the computer from the human, the computer passes the Turing Test, otherwise the computer fails.

A. Searle's "Chinese room"

Strong AI: the "appropriately programmed computer *is* a mind, in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states."

(RR, 384) According to strong AI:

- i) The machine literally understands stories and the like.
- ii) what the machine and its programs do explains the human ability to understand a story and answer questions about it. (RR 385)

Roughly: Searle describes himself locked in a room, given a lot of Chinese writing and rules for producing new Chinese writing to give back. Does Searle understand Chinese?

He claims that he doesn't understand Chinese and there is no reason to think that what computers do provides even part of an explanation of how we understand.

*Objections:*

1. Systems Reply

--but how could an individual together with bits of paper understand when the individual doesn't?

--all sorts of non-cognitive systems will end up counting as cognitive.

--cognition is not simply a matter of "formal" symbol manipulation.

2. Robot Reply

--Searle could be the homunculus in a robot, but still doesn't understand.

3. Brain Simulator Reply

--why do we need to know how the brain works if functionalism allows multiple realizability?

--a complicated brain-like system still wouldn't understand (water pipe example).

4. Combination Reply

--what help does adding all (1-3) together do?

5. Other Minds Reply

--misses the point: we're not worrying about how we know of others mental lives, but what it is to have mental states at all.

6. Many Mansions Reply

---this is just wishful thinking.

Searle's view:

In considering whether a machine could think, Searle replies, "My own view is that *only* a machine could think, and indeed only very special kinds of machines, namely brains and machines that had the same causal powers as brains...Whatever else intentionality is, it is a biological phenomenon, and it is as likely to be causally dependent on the specific biochemistry of its origins as lactation, photosynthesis, or any other biological phenomena." (RR 396).

*B. Lycan's reply.*

Let's grant that:

Intelligence is the ability to balance a wide variety of goals under a wide variety of outside circumstances.

Modern-day computers lack intelligence for two reasons:

- They rely on humans to feed them information, which greatly limits their flexibility.
- They can't develop their own goals.

But if Lycan's definition captures what we mean by intelligence, then maybe the Turing test doesn't check for intelligence.

Perhaps: Computers with the right programs are literally capable of thinking and understanding, but what counts as the "right programs" isn't determined by the Turing Test. If computers could be developed so that they are more flexible and set their own goals, does it matter whether they are made of metal and circuits?

*C. Jackson's "Knowledge Argument"*

Some have argued that functionalism cannot accommodate the "felt quality" of experience, or the qualitative aspect of perception. Consider Mary from Jackson's essay:

- (1) Before her release, Mary knows all the physical facts.
- (2) Then Mary learns a *new* fact: what yellow experience is like.
- (3) Facts about what experience is like are not physical facts.

Does this provide an argument against functionalism?