5. An argument against dualism and the Turing test

Martín Abreu Zavaleta

June 3, 2014

1 The argument from overdetermination

Some people have used this argument against dualism. Let's start by introducing a couple of definitions:

Causal overdetermination: An event x is causally overdetermined just in case it is caused by at least two different events, where each of these events would have sufficed by itself to cause x.

Causal overdetermination is usually illustrated with the case of the firing squad: suppose a firing squad is instructed to kill a given person, A. The shooters fire in such a way that multiple bullets hit A *at the same time*, each of them reaching a vital spot, the destruction of which would cause immediate death. Had only one of the bullets reached its target, it would have sufficed to cause A's death, but as it happens, all the bullets reached their target, and so, they all caused A's death. In this case, we say that A's death is causally overdetermined. **Question:** can you think of other cases of causal overdetermination? Should we expect causal overdetermination to be a widespread phenomenon?

Now let's introduce the thesis of the *causal closure of the physical domain:*

Causal closure: If a *physical* event has a cause, it has a sufficient *physical* cause.

According to this claim, if we trace the chains of causes of a physical event, at every step in the chain, we will find a physical event. For instance, suppose that we are playing pool, and I put ball 8 in one of the corner pockets. We may ask: what caused ball 8 to go into that pocket? In this case, let's suppose, what caused that event was the event of the cue ball hitting it in such and such way. What caused that event was my hitting the cue ball with the cue, and what caused this was the movement of my arms in such and such ways while holding the cue. We could trace the event of ball 8 going into the pocket up to the big bang! What matters is that all the events in this chain had *physical causes*, if they had a cause at all. Part of the reason to accept this claim is that we don't really have any reason to posit any non-physical causes in order to explain physical events.

Now we are in a position to state the argument. We can start with the premise that some mental events cause physical events. So far, we are not making any assumptions whether physicalism or dualism is correct, we are just stating the intuitively compelling claim that things like our beliefs, desires and sensations cause some physical events: if I feel like my hand is burning, I will remove it from the fire; if I believe that there is chocolate in the cupboard and I want chocolate, I will take a look at the cupboard, etc.

By *causal closure*, all those physical events—my removing my hand from the fire, or my moving towards the cupboard—have physical causes. This is where the argument gets interesting. Now we have two options, either:

(i) The beliefs, desires and sensations that caused the events in question are themselves physical in nature, or

(ii) The beliefs, desires and sensations that caused the events in question are *not physical* in nature.

If the first is true, then we have already accepted physicalism, and our work is done. We can make sense of the fact that our believes, desires and sensations cause our movements and so on, because those believes, desires and sensations are themselves physical.

Now let's take a look at the second claim. If (ii) was the case, then every physical event which is allegedly caused by a mental event (our having a belief, desire, sensation, etc.) would be *causally overdetermined*. The reason is that every such physical event e would have a physical cause—by the *causal closure of the physical domain*—but if our beliefs, desires and sensations also cause e, and are not themselves physical, this means that all such events e have two different causes: one mental and one physical.

Now, there seems to be no problem with some event or other being causally overdetermined, but by the argument above, *interactionist dualism* is committed to the claim that *every physical event that has a mental cause is causally overdetermined.* This seems hard to believe: if interactionist dualism is correct, then there would be widespread and systematic overdetermination of a lot of physical events. Why should we think that such overdetermination obtains? And what if the mental and the physical events go out of sync? A simpler theory is just the theory that no such overdetermination occurs, and (i) is true, but that is physicalism.

Notice that the argument from overdetermination doesn't attempt against all kinds of dualism. In particular, the following version of dualism can avoid the objection from causal overdetermination:

Epiphenomenalism: Mental events are not the same as physical events, and mental events *do not cause* physical events.

Since epihphenomenalism claims that mental events don't cause physical events, it is not committed to the widespread overdetermination of certain kinds of physical events. However, it avoids the objection at great cost: it's hard to believe that our mental states have absolutely no influence on the physical world.

So far we have examined very general views about the nature of the mind, but there are also more detailed views. We'll start considering some of them, but first it will be useful to take a look at a test offered by Alan Turing.

2 Turing's test

We have been talking a lot about mentality, mental states, and whether mental things are physical or not. However, we haven't yet wondered how we can find out whether something has mentality or not; that is, whether something has mental states, or is intelligent, or something like that.

In his paper "Computer Machinery and Intelligence", Alan Turing offers a test that, according to him, will allow us to determine whether a computer is genuinely capable of thinking and genuinely intelligent. The test works like this:

> We have two rooms. In one, there is a human, the interrogator. In the other room, there is another human and a computer. The computer and the other person are only known as X and Y to the interrogator, and she is not allowed to see them. The interrogator's objective is to find out which of the two is the human and which is the computer. In order to do this, she may ask any question she wants to the other participants by means of a keyboard, and see their answers by means of a screen. The objective of the computer is to trick the interrogator into thinking that it is the human, whereas the other person's job is to help the interrogator.

If the computer wins the game often enough, then it passes the test, and so it is genuinely capable of thinking and genuinely intelligent. **Question:** What do you think about this test? If a computer passes this test, what would that entail?

Turing examines various objections to his test, of which we'll examine two.

Machines don't make mistakes

Turing states the objection as follows:

> It is claimed that the interrogator could distinguish the machine from the man simply by setting them a number of problems in arithmetic. The machine would be unmasked because of its deadly accuracy.

In order to see the import of the objection, it is important to distinguish between two kinds of mistakes, what Turing calls *errors of functioning* and *errors of conclusion*:

> Errors of functioning are due to some mechanical or electrical fault which causes the machine to behave otherwise than it was designed to do. In philosophical discussions oe likes to ignore the possibility of such errors [...] Errors of conclusion can only arise when some meaning is attached to the output signals from the machine. The machine might, for instance, type out mathematical equations, or sentences in English. When a false proposition is typed we say that the machine has committed an error of conclusion.

Turing points out that there is no reason to believe that the machine won't make the second kind of mistake, and errors of the first kind are irrelevant. It may be, for instance, that the machine knowingly types a false claim in response to one of the questions of the interrogator, but it may also be that it makes a mistake in a more natural sense.

It all depends on how the machine is programmed. If the machine is programmed in the way in which a calculator is, then it probably won't make mistakes. But if, on the other hand, it is programmed to follow the kind of reasonings that we use when we perform arithmetic operations, then it could be as prone to error as we are.**Question:** What other things are there that you think machines can't do, even in principle?

Lady Lovelace's objection

Lady Lovelace offered our most detailed information of Babage's analytical engine, one of the first computers. She writes:

> The Analytical Engine has no pretensions to originate anything. It can do *whatever we know how to order it to perform*

The objection is that a machine can't really do anything new, and can only have a fixed pattern of behavior.

Is this true? In a way, it is true that a machine can only do what its program tells it to do. However, this doesn't mean that the computer will always do the same things, nor does it mean that anyone who knows how it is programmed will thereby know what the machine will do next, at any given time.

For instance, we can program machines that "learn". Jim Pryor offers the following example:

> Computers can also be programmed to revise their own programs, to "learn from their mistakes." For instance, a chess-playing computer can be programmed to thoroughly analyze its opponents' strategy whenever it loses a game, and incorporate those winning strategies into its own database. In this way, it can learn how to use those strategies in its own games, and learn how to look out for them in its opponents' games. This kind of computer would get better and better every time it played chess. It would quickly become able to do things its programmers never anticipated (though of course they're the ones who programmed it to learn from its mistakes in this way).(http://goo.gl/ZMjqZm)

Attitudes towards the Turing test

There are at least three attitudes that we can take towards Turing's claims that passing the test is sufficient for being intelligent or having thoughts:

1. Completely deny that passing the Turing test guarantees being able to think or being intelligent. At best, the Turing test is good for *simulating* thought and intelligence, not for *genuinely* having thoughts and intelligence.

2. Think that passing the Turing test gives us good reason to believe that a computer can think or is intelligent, but doesn't guarantee being intelligent or having thoughts.

3. Think that passing the Turing test *suffices* for being intelligent or having thoughts. According to this response, being intelligent *is just* to behave in certain ways. In particular, being intelligent *is just* to behave in the ways that something would need to behave in so that it passes Turing's test.