

22. The knowledge argument

Martín Abreu Zavaleta

July 1st, 2014

1 The argument

Jackson's argument can be reconstructed like this:

- (1) At time t_0 , Mary has all the physical information concerning human color vision.
- (2) At time t_1 , when Mary is released into the world, she comes to have information she didn't have before.
- (3) Therefore, this new information is not physical information.

Perhaps the most explicit presentation of the argument in Jackson's paper is this:

What will happen when Mary is released from her black and white room or is given a color television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then is it inescapable that her previous knowledge was incomplete. But she had all the physical information. Ergo there is more to have than that, and Physicalism is false.

The general idea is that someone may know all the physical information and still be missing some piece of information. Presumably, the information we would be missing would be about how something looks, or *how it is like* to look at that thing. We'll see later how to refine the argument, but for now this will do. Now let's look at ways of distinguishing this argument from others.

If we see the argument in this last way, we can understand the examples about Mary and Fred as trying to produce the intuition that someone may have all the physical information and still lack some information.

The modal argument

Here's the modal argument, reconstructed by Jackson:

- (1) No amount of physical information about someone *logically* entails that she has conscious experiences.
- (2) Thus, it is possible that there are things physically identical to us, but without conscious experiences; therefore,
- (3) Conscious experiences are not physical things (since the consciousless things are physically identical to us).

How is this different from the knowledge argument? Jackson claims that it rests on a different intuition, namely, the intuition that it is possible for there to be something physically identical to us, but without conscious experience. He claims that the knowledge argument doesn't rest on this intuition. Instead, it's supposed to rest on the intuition that there is something Mary didn't know even if she had all the physical information. **Question:** Jackson thinks that his argument is different from Nagel's in "What is it like to be a bat?" do you think this is right?

Epiphenomenalism

Jackson thinks that qualia, or phenomenal properties of mental states are causally inefficacious: their possession or absence makes no difference to the physical world. However, he thinks that the instantiation of qualia may make a difference to other mental states. Most of his defense will consist in rejecting some reasons to think that qualia are causally efficacious.

- (i) It seems obvious that the way pain feels is partly responsible for our reactions to pain, so the way pain feels must have some causal powers. Jackson merely points out that just because two phenomena are correlated, that doesn't mean that one causes the other. In particular, if there is some more plausible theory (e.g. one that explains how the two phenomena are caused by a third thing, but independent of each other), we should reject that the way pain feels has any causal powers.
- (ii) Some people claim that we may assume that qualia evolved over time, and so we may expect qualia to be conducive to survival. After all, according to natural selection the traits that evolve over time are the ones that are conducive to survival. Jackson points out that not every trait has to be selected. It may be that some trait is not conducive to survival, but is a by-product of something that conduces to survival. If this is the case, evolutionary theory is compatible with epiphenomenalism.
- (iii) Our way of knowing that other people have qualia depends on our observation of their behavior. If qualia were not causally efficacious, observation of people's behavior wouldn't give us much reason to think that they have qualia. So qualia must be causally efficacious. Jackson continues his strategy and points out that behavior may still give us evidence for the presence of qualia in other people, if only because we may postulate that whatever causes the behavior causes the qualia too.
- (iv) If qualia weren't causally efficacious, they would play no explanatory role. Our best theory would not need to posit them. Jackson acknowledges that qualia may not play any interesting explanatory role, but claims that this doesn't show that there are no qualia.

2 Refinements and observations

The first thing to note is that, the way Jackson originally stated the argument, there is an ambiguity in the notion of information. In one sense, 'Hesperus is a planet' and 'Phosphorus is a planet' transmit different information, since someone may know that Hesperus is a planet but still not know that Phosphorus is a planet. In the other sense, 'Hesperus is a planet' and 'Phosphorus is a planet' give us the same information, since they express the same proposition, or are true under exactly

the same circumstances. Which sense of information is the one required by the argument? Let's examine both.

For the sake of clarity, let's call the first notion of information *narrow* and the second *wide*.

With narrow information, we may present the argument as follows:

- 1 At time t_0 , for every truth p about color vision described using only physical vocabulary, Mary knew that p .
- 2 Before Mary comes out of her lab, there is some $q \neq p$ such that Mary didn't know that q .
- 3 Therefore, not all knowledge is physical knowledge.

Here the conclusion is to be understood in the following terms: that one can know something when described in a particular way, but not under a different way. This is just like knowing that Hesperus is a planet but not knowing that Phosphorus is a planet.

This sort of conclusion need not worry the physicalist. Physicalism is usually meant as a metaphysical thesis about the nature of the mental. In its simpler formulation, it claims that mental things *are* physical things, or reducible to physical things. This metaphysical claim is compatible with the claim that mental vocabulary may just not be translatable to physical vocabulary, or that mental concepts are not reducible to physical concepts.

However, there is a stronger version of the argument, using the wide notion of information. Jackson's own presentation is as follows:

- (1) Mary (before her release) knows everything physical there is to know about other people.
- (2) Mary (before her release) does not know everything there is to know about other people (because she learns something about them on her release); therefore
- (3) There are truths about other people (and herself) which escape the physicalist story.

On this way of understanding the argument, what Mary doesn't know is a *fact*. It's not only that she lacks the appropriate concepts, or that she doesn't know under a particular description. What she fails to know is a fact, and given that she knew all the physical facts, this new fact can't be physical.

Questions: What do you think of this version of the argument? Has Jackson made a compelling case that the main premise is true?