

Handout 10: the Knowledge argument against physicalism

"Mary is a brilliant scientist who is...forced to investigate the world from a black and white room *via* a black and white television monitor". She "acquires, let us suppose, all the physical information there is to obtain when we see ripe tomatoes...use terms like 'red'...and so on...What will happen when Mary is released from her black and white room?...Will she *learn* anything or not?"

"It just seems obvious that she will learn something about the world and our experience of it...But she had all the physical information. Ergo...Physicalism is false"

1. Imprisoned Mary knows all the physical facts.

Hence:

2. If physicalism is true, Mary (before her release) knows, or can know, all the facts.

3. After her release, Mary *learns* something -- something she couldn't have known while imprisoned.

4. If Mary learns something, she learns a *fact*.

Hence (from 3, 4):

5. Mary learns a fact.

Hence (from 2, 5):

6. Physicalism is false.

Jackson draws a further conclusion:

7. Our experiences have "qualia" ("The whole thrust of the knowledge argument is that Mary...does not know about certain qualia..." ("What Mary didn't know", p. 279)). Qualia are "certain [nonphysical] features of bodily sensations [and] perceptual experiences... the hurtfulness of pains, the itchiness of itches [etc.]" (p. 273).

(Rough) gloss on the move from 1 to 2: not knowing something is not being able to decide between rival possibilities -- in other words not being able to tell which possible world one is in; so if imprisoned Mary doesn't know something then she can't tell exactly which possible world she is in; but if physicalism is true she plainly *can* decide what possible world she is in, because if physicalism is true then the totality of the physical facts eliminates all possibilities but one, and she knows all the physical facts.^[1] Therefore if

physicalism is true Mary can decide between any rival possibilities, so she knows (or can know) everything.

Four responses

David Lewis denies premise 4. Daniel Dennett denies premise 3. Brian Loar ("Phenomenal states") thinks 2 does not follow from 1. Gilbert Harman (in a paper called "The intrinsic quality of experience", not reprinted in Chalmers) denies premise 1 (or seems to, at any rate -- see below).

Dennett denies premise 3 because he thinks that Mary's situation is not easily imaginable: "It is of course true that in any realistic, readily imaginable version of the story, Mary would come to learn something, but in any realistic, readily imaginable version she might know a lot, but she would not know everything physical" (from *Consciousness Explained*).

Lewis, on the other hand, thinks Mary *learns*, but that she doesn't learn a *fact*: rather, she acquires "know-how", and thereby an *ability* (an ability to remember, imagine and recognize the experience of seeing something red). (See "What experience teaches".) And since physicalism does not imply that imprisoned Mary must be capable of acquiring every *ability*, one cannot counter Lewis's reply by simply reformulating premise 2 to concern abilities rather than knowledge. For some objections to Lewis, see Loar, pp. 303-4.

Perhaps the most popular response is Loar's: resist the move from 1 to 2, with the following motivation (see p. 464). All god has to do to make water come out of faucets is to make H₂O come out of faucets (cf. the physicalist claim that all god has to do to make pain is to make c-fibers fire). In other words, every possible world in which H₂O comes out of faucets is a world in which water comes out of taps. In yet other words: 'if H₂O comes out of faucets then water comes out of faucets' is a necessary truth. But it does not follow that if superchemist Sally were locked in a room and knew that H₂O comes out of faucets then she would need no more empirical information in order to know that *water* comes out of faucets. The knowledge argument (so this response goes) illicitly draws an *epistemological* conclusion (that black-and-white Mary can know everything) from a *metaphysical* thesis (that the physical facts determine all the facts). For objections to this response, see Chalmers, "Consciousness and...", pp. 253-7

According to Harman, "The person blind from birth fails to know what it is like it to see something red because he or she does not fully understand what it is for something to be red, that is, because he or she does not have the full concept of something's being red. So...the person blind from birth does not know all the functional facts..." (from "The intrinsic quality of experience" -- copies of this paper are available on request). Thus, if we run a revised version of the knowledge argument with Mary being *blind from birth* and with the first premise being "Imprisoned Mary knows all the physical *and functional* facts", then Harman would deny the first premise. (More explicitly, he would claim that if we stipulate that Mary is blind from birth then she cannot be omniscient with respect to

the physical and functional facts -- the description of Mary's initial predicament is inconsistent.) Exercise: Suppose Harman's response works against this revised version of the knowledge argument. Would it still work against the the knowledge argument as presented above?

If you want to know more about the current state of play, see the instructor's [review](#) of [*There's something about Mary*](#)

[1] This isn't quite right. If physicalism is true, the totality of physical facts doesn't eliminate *all* possibilities but one: a possible world exactly like our own physically but with the addition of some epiphenomenal angels or "spooks" won't get eliminated (see Lewis, pp. 286-7). But, as Lewis explains, this wrinkle doesn't matter for present purposes. (This footnote is for enthusiasts only.)

MIT OpenCourseWare
<http://ocw.mit.edu>

24.09 Minds and Machines
Fall 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.