

## COMMENTS AND CRITICISM

## WHAT MARY DIDN'T KNOW\*

MARY is confined to a black-and-white room, is educated through black-and-white books and through lectures relayed on black-and-white television. In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in *completed* physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles. If physicalism is true, she knows all there is to know. For to suppose otherwise is to suppose that there is more to know than every physical fact, and that is just what physicalism denies.

Physicalism is not the noncontroversial thesis that the actual world is largely physical, but the challenging thesis that it is entirely physical. This is why physicalists must hold that complete physical knowledge is complete knowledge simpliciter. For suppose it is not complete: then our world must differ from a world,  $W(P)$ , for which it is complete, and the difference must be in nonphysical facts; for our world and  $W(P)$  agree in all matters physical. Hence, physicalism would be false at our world [though contingently so, for it would be true at  $W(P)$ ].<sup>1</sup>

It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red, say. This is rightly described as *learning*—she will not say “ho, hum.” Hence, physicalism is false. This is the knowledge argument against physicalism in one of its manifestations.<sup>2</sup> This note is a reply to three objections to it mounted by Paul M. Churchland.<sup>†</sup>

\* I am much indebted to discussions with David Lewis and with Robert Pargetter.

<sup>1</sup> The claim here is not that, if physicalism is true, only what is expressed in explicitly physical language is an item of knowledge. It is that, if physicalism is true, then if you know everything expressed or expressible in explicitly physical language, you know everything. *Pace* Terence Horgan, “Jackson on Physical Information and Qualia,” *Philosophical Quarterly*, xxxiv, 135 (April 1984): 147–152.

<sup>2</sup> Namely, that in my “Epiphenomenal Qualia,” *ibid.*, xxxii, 127 (April 1982): 127–136. See also Thomas Nagel, “What Is It Like to Be a Bat?,” *Philosophical Review*, LXXXIII, 4 (October 1974): 435–450, and Howard Robinson, *Matter and Sense* (New York: Cambridge, 1982).

<sup>†</sup> “Reduction, Qualia, and the Direct Introspection of Brain States,” this JOURNAL, LXXXII, 1 (January 1985): 8–28. Unless otherwise stated, future page references are to this paper.

## I. THREE CLARIFICATIONS

The knowledge argument does not rest on the dubious claim that logically you cannot imagine what sensing red is like unless you have sensed red. Powers of imagination are not to the point. The contention about Mary is not that, despite her fantastic grasp of neurophysiology and everything else physical, she *could not imagine* what it is like to sense red; it is that, as a matter of fact, she *would not know*. But if physicalism is true, she would know; and no great powers of imagination would be called for. Imagination is a faculty that those who *lack* knowledge need to fall back on.

Secondly, the intensionality of knowledge is not to the point. The argument does not rest on assuming falsely that, if *S* knows that *a* is *F* and if *a* = *b*, then *S* knows that *b* is *F*. It is concerned with the nature of Mary's total body of knowledge before she is released: is it complete, or do some facts escape it? What is to the point is that *S* may know that *a* is *F* and *know* that *a* = *b*, yet arguably not know that *b* is *F*, by virtue of not being sufficiently logically alert to follow the consequences through. If Mary's lack of knowledge were at all like this, there would be no threat to physicalism in it. But it is very hard to believe that her lack of knowledge could be remedied merely by her explicitly following through enough logical consequences of her vast physical knowledge. Endowing her with great logical acumen and persistence is not in itself enough to fill in the gaps in her knowledge. On being let out, she will not say "I could have worked all this out before by making some more purely logical inferences."

Thirdly, the knowledge Mary lacked which is of particular point for the knowledge argument against physicalism is *knowledge about the experiences of others*, not about her own. When she is let out, she has new experiences, color experiences she has never had before. It is not, therefore, an objection to physicalism that she learns *something* on being let out. Before she was let out, she could not have known facts about her experience of red, for there were no such facts to know. That physicalist and nonphysicalist alike can agree on. After she is let out, things change; and physicalism can happily admit that she learns this; after all, some physical things will change, for instance, her brain states and their functional roles. The trouble for physicalism is that, after Mary sees her first ripe tomato, she will realize how impoverished her conception of the mental life of *others* has been *all along*. She will realize that there was, all the time she was carrying out her laborious investigations into the neurophysiologies of others and into the functional roles of their internal states, something about these people she was quite unaware of. All along their experiences (or many of them, those got from tomatoes, the

sky, . . .) had a feature conspicuous to them but until now hidden from her (in fact, not in logic). But she knew all the physical facts about them all along; hence, what she did not know until her release is not a physical fact about their experiences. But it is a fact about them. That is the trouble for physicalism.

## II. CHURCHLAND'S THREE OBJECTIONS

(i) Churchland's first objection is that the knowledge argument contains a defect that "is simplicity itself" (23). The argument equivocates on the sense of 'knows about'. How so? Churchland suggests that the following is "a conveniently tightened version" of the knowledge argument:

- (1) Mary knows everything there is to know about brain states and their properties.
  - (2) It is not the case that Mary knows everything there is to know about sensations and their properties.
- Therefore, by Leibniz's law,
- (3) Sensations and their properties  $\neq$  brain states and their properties (23).

Churchland observes, plausibly enough, that the type or kind of knowledge involved in premise 1 is distinct from the kind of knowledge involved in premise 2. We might follow his lead and tag the first 'knowledge by description', and the second 'knowledge by acquaintance'; but, whatever the tags, he is right that the displayed argument involves a highly dubious use of Leibniz's law.

My reply is that the displayed argument may be convenient, but it is not accurate. It is not the knowledge argument. Take, for instance, premise 1. The whole thrust of the knowledge argument is that Mary (before her release) does *not* know everything there is to know about brain states and their properties, because she does not know about certain qualia associated with them. What is complete, according to the argument, is her knowledge of matters physical. A convenient and accurate way of displaying the argument is:

- (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.

What is immediately to the point is not the kind, manner, or type of knowledge Mary has, but *what* she knows. What she knows be-

forehand is *ex hypothesi* everything physical there is to know, but is it everything there is to know? That is the crucial question.

There is, though, a relevant challenge involving questions about kinds of knowledge. It concerns the *support* for premise 2'. The case for premise 2' is that Mary learns something on her release, she acquires knowledge, and that entails that her knowledge beforehand (*what* she knew, never mind whether by description, acquaintance, or whatever) was incomplete. The challenge, mounted by David Lewis and Laurence Nemirow, is that on her release Mary does *not* learn something or acquire knowledge in the relevant sense. What Mary acquires when she is released is a certain representational or imaginative ability; it is knowledge *how* rather than knowledge *that*. Hence, a physicalist can admit that Mary acquires something very significant of a knowledge kind—which can hardly be denied—without admitting that this shows that her earlier factual knowledge is defective. She knew *all that* there was to know about the experiences of others beforehand, but lacked an ability until after her release.<sup>3</sup>

Now it is certainly true that Mary will acquire abilities of various kinds after her release. She will, for instance, be able to imagine what seeing red is like, be able to remember what it is like, and be able to understand why her friends regarded her as so deprived (something which, until her release, had always mystified her). But is it plausible that that is *all* she will acquire? Suppose she received a lecture on skepticism about other minds while she was incarcerated. On her release she sees a ripe tomato in normal conditions, and so has a sensation of red. Her first reaction is to say that she now knows more about the kind of experiences others have when looking at ripe tomatoes. She then remembers the lecture and starts to worry. Does she really know more about what their experiences are like, or is she indulging in a wild generalization from one case? In the end she decides she does know, and that skepticism is mistaken (even if, like so many of us, she is not sure how to demonstrate its errors). What was she to-ing and fro-ing about—her abilities? Surely not; her representational abilities were a known constant throughout. What else then was she agonizing about than whether or not she had gained factual knowledge of others? There would be nothing to agonize about if ability was *all* she acquired on her release.

<sup>3</sup> See Laurence Nemirow, review of Thomas Nagel, *Mortal Questions*, *Philosophical Review*, LXXXIX, 3 (July 1980): 473–477, and David Lewis, "Postscript to 'Mad Pain and Martian Pain'," *Philosophical Papers*, vol. 1 (New York: Oxford, 1983). Churchland mentions both Nemirow and Lewis, and it may be that he intended his objection to be essentially the one I have just given. However, he says quite explicitly (bottom of p. 23) that his objection does not need an "ability" analysis of the relevant knowledge.

I grant that I have no *proof* that Mary acquires on her release, as well as abilities, factual knowledge about the experiences of others—and not just because I have no disproof of skepticism. My claim is that the knowledge argument is a valid argument from highly plausible, though admittedly not demonstrable, premises to the conclusion that physicalism is false. And that, after all, is about as good an objection as one could expect in this area of philosophy.

(ii) Churchland's second objection (24/5) is that there must be something wrong with the argument, for it proves too much. Suppose Mary received a special series of lectures over her black-and-white television from a full-blown dualist, explaining the "laws" governing the behavior of "ectoplasm" and telling her about qualia. This would not affect the plausibility of the claim that on her release she learns something. So if the argument works against physicalism, it works against dualism too.

My reply is that lectures about qualia over black-and-white television do not tell Mary all there is to know about qualia. They may tell her some things about qualia, for instance, that they do not appear in the physicalist's story, and that the quale we use 'yellow' for is nearly as different from the one we use 'blue' for as is white from black. But why should it be supposed that they tell her everything about qualia? On the other hand, it is plausible that lectures over black-and-white television might in principle tell Mary everything in the physicalist's story. You do not need color television to learn physics or functionalist psychology. To obtain a good argument against dualism (attribute dualism; ectoplasm is a bit of fun), the premise in the knowledge argument that Mary has the full story according to physicalism before her release, has to be replaced by a premise that she has the full story according to dualism. The former is plausible; the latter is not. Hence, there is no "parity of reasons" trouble for dualists who use the knowledge argument.

(iii) Churchland's third objection is that the knowledge argument claims "that Mary could not even *imagine* what the relevant experience would be like, despite her exhaustive neuroscientific knowledge, and hence must still be missing certain crucial information" (25), a claim he goes on to argue against.

But, as we emphasized earlier, the knowledge argument claims that Mary would not know what the relevant experience is like. What she could imagine is another matter. If her knowledge is defective, despite being all there is to know according to physicalism, then physicalism is false, whatever her powers of imagination.

FRANK JACKSON

Monash University