

Doubts About the Knowledge Argument

Michael Pelczar

The Knowledge Argument purports to disprove physicalism, the view that conscious experience in all its details is in some sense automatically given with (identical to, entailed by, supervenient upon, grounded in) purely physical features of reality. It goes like this:¹

First premise:

You could know the whole physical story about a creature that has certain conscious experiences, without knowing what it is like to have such experiences.

Second premise:

If you could know the whole physical story about a creature that has certain conscious experiences, without knowing what it is like to have such experiences, then there's more to having such experiences than having a certain kind of physical nature.

Conclusion:

So there's more to having certain conscious experiences than having a certain kind of physical nature.

¹The clearest and most influential presentation of the Knowledge Argument is the one found in (Jackson, 1982). The literature surrounding the argument is vast.

By “someone who knows the whole physical story about some creature,” we mean someone who knows, for every physical feature of that creature, that the creature has that feature. And by a “physical feature” of something, we mean a feature (part, property, process, state, etc.) of it that is, in theory, completely describable in terms of basic physics. So, for example, if you knew the whole physical story about the Moon, you would know, for every feature of the Moon that is completely describable in terms of basic physics, that the Moon had that feature; and, if you knew the whole physical story about *me*, you would know, for every feature of mine that is completely describable in terms of basic physics, that I had that feature.

The first premise of the Knowledge Argument receives support from thought-experiments in which, for example, someone learns everything about the physical nature of a bat, without learning what it is like to have the sort of experience that a bat has when using its sense of echolocation to navigate through its environment, or in which someone confined to a grayscale facility learns everything about the physical nature of normally-situated human beings, without learning what it is like to have the sort of experiences that those human beings have when they see red tomatoes, yellow bananas, etc. These thought-experiments certainly seem to point to genuine logical possibilities, or even to genuine empirical possibilities.

In any event, I am happy to grant the first premise of the Knowledge Argument. My dispute is with the second premise.

The second premise of the argument says, in effect, that a certain combination of states of affairs is impossible. It says that there can't be

a situation in which (a) you know the whole physical story about some sentient creature; (b) you don't know what it's like to have certain experiences that that creature has; and, (c) there's nothing more to that creature's having those experiences than that the creature has a certain physical nature. In terms of the famous Mary thought-experiment, the second premise says that the following statements constitute an inconsistent triad:²

- (1) Mary knows the whole physical story about some person, Jack, who has reddish experience.
- (2) Mary doesn't know what it's like to have reddish experience.
- (3) There is nothing more to Jack's having reddish experience than that he has a certain physical nature.

Here, "reddish experience" means the kind of experience normal human beings normally have when viewing red objects under normal lighting conditions, and "there is nothing more to Jack's having reddish experience than that he has a certain physical nature" means that it is a necessary consequence of the fact that Jack has certain physical features that he has reddish experience.

To cast doubt on the second premise of the Knowledge Argument, it is sufficient to show that (1), (2), and (3) can all be true together, and the key to showing this is to scrutinize the meaning of statement (2).

What do we mean when we say that someone knows what it is like to have reddish experience?

²For the Mary thought-experiment, see (Jackson, 1982, 130).

Let's try to answer that question by asking another question. Supposing that someone has never had reddish experience and is not having reddish experience now, are we prepared to say that this person knows what it's like to have reddish experience? Can we even conceive of a situation in which a person who never has reddish experience is correctly describable as knowing what it's like to have reddish experience?

Actually, it seems we can. At least, we can conceive of a situation in which some atoms spontaneously jump together to form a perfect living duplicate of me at a moment when I am staring into a cloudless blue sky. Unlike me, my duplicate has never had reddish experience (nor is he having reddish experience now). Still, it seems rather odd to deny that my duplicate knows what it's like to have reddish experience, given that I myself know what this is like, and given that he resembles me as perfectly as he does. Perhaps there are some who will insist that despite this resemblance, my duplicate does not know what it is like to have reddish experience. If they are correct to insist on this, it makes my criticism of the Knowledge Argument even more straightforward. But, to err on the side of caution, let's suppose that my duplicate *does* know what it's like to have reddish experience, despite the fact that he has never had such experience.

Returning to the question of what we mean when we say things like, "Jack knows what it's like to have reddish experience," let's ask: if someone has never had reddish experience, is not having reddish experience now, and does not resemble anyone who has had reddish experience in the relevant respects in which my duplicate in the case described above resembles me, are we prepared to say that this person knows what it's

like to have reddish experience? Can we even conceive of a situation in which such a person is correctly describable as knowing what it's like to have reddish experience?

I submit that we cannot. More generally, I submit that part of what we mean by saying that someone knows what it's like to have an experience with conscious quality *Q* is that the person is having an experience with quality *Q*, or has had an experience with quality *Q*, or resembles a person who has had an experience with quality *Q* in all relevant respects other than in respect to having actually had an experience with quality *Q* (*à la* duplicate of me in the case described above). (By saying that this is part of what we mean by saying that someone knows what it's like to have an experience with *Q*, I mean that when we say this, we mean to express a proposition that we wouldn't accept if we thought that the stated disjunctive condition were false.)

Maybe there is someone who would dispute this claim. For present purposes, that doesn't matter. It doesn't matter, because a proponent of the Knowledge Argument is certainly not in a position to dispute the claim. Why not? Because if the claim is false, then we should not accept the Knowledge Argument's first premise. If there is a way to know what it is like to have reddish experience without ever having reddish experience (and without relevantly resembling anyone who has had reddish experience in the manner described above), then we cannot rule out the possibility that anyone who knew the whole physical story about Jack would thereby know what it was like to have reddish experience, in whatever way one can (allegedly) know this without ever having reddish experience (or relevantly resembling someone who has). But if we can-

not rule out that possibility, then we should not accept the first premise of the Knowledge Argument.

So a proponent of the Knowledge Argument has no choice but to agree that part of what it means to say that someone knows what it's like to have reddish experience is that he or she is having reddish experience or has had reddish experience or relevantly resembles someone who has had reddish experience.

Let's go back to the putatively inconsistent triad given above. The second statement of the triad says that Mary doesn't know what it's like to have reddish experience. As we have seen, this statement is true as long as it is true that Mary is not having reddish experience and has never had reddish experience and does not relevantly resemble anyone who has had reddish experience. And this latter statement ("Mary is not having reddish experience and has never . . . etc.") is indeed true, given the details of the thought-experiment.

Now let's suppose that the first statement of the triad is true; i.e., let's suppose that Mary knows the whole physical story about Jack (who has reddish experience). Certainly a proponent of the Knowledge Argument must allow that this is compatible with Mary's never having reddish experience or relevantly resembling anyone who has; otherwise, the whole Mary thought-experiment would have to be rejected as incoherent. So we can safely say that (1) is compatible with (2).

But is the conjunction of (1) and (2) compatible with (3)? Well, the conjunction is true provided that (i) Mary knows the whole physical story about Jack, and, (ii) Mary is not having, and has never had, and does not relevantly resemble anyone who has had, reddish experience.

But (i) and (ii) are jointly compatible with the claim that there's no more to having reddish experience than having a certain physical nature. There is no contradiction in saying: "Mary knows the whole physical story about Jack, and Mary has never had (or relevantly resembled anyone who has had) reddish experience, and Jack's reddish experiences are purely physical brain-processes." Or if there is a contradiction here, it is highly non-obvious, and would have to be established by some convincing independent argument.

A physicalist can grant that Mary knows the whole physical story about Jack, and grant that Mary never has reddish experience (and never relevantly resembles anyone who has had reddish experience), while insisting that having a reddish experience is just a matter of satisfying certain purely physical conditions—physical conditions that Jack satisfies, and Mary doesn't. At least, proponents of the Knowledge Argument have said nothing to persuade us otherwise.

So: we have at present no reason to regard the allegedly inconsistent triad given above—(1), (2), (3)—as genuinely inconsistent. But the second premise of the Knowledge Argument is equivalent to the claim that this triad (or any triad of the same form) *is* inconsistent. I conclude that we have at present no reason to accept the second premise of the Knowledge Argument.

References

Jackson, Frank. 1982. Epiphenomenal qualia. *The Philosophical Quarterly*, **32**(127), 23–36.