The Knowledge Argument

(penultimate draft)

In *The Encyclopedia of Philosophy*, 2005 (Macmillan)

Brie Gertler
University of Virginia

The definitive statement of the Knowledge Argument was formulated by Frank Jackson, in a paper entitled "Epiphenomenal Qualia" that appeared in The Philosophical Quarterly in 1982. Arguments in the same spirit had appeared earlier (Broad 1925, Robinson 1982), but Jackson's argument is most often compared with Thomas Nagel's argument in "What is it Like to be a Bat?" (1974). Jackson, however, takes pains to distinguish his argument from Nagel's. This entry will follow standard practice in focusing on Jackson's argument, though I will also describe the main points of alleged similarity and dissimilarity between these two arguments.

The Knowledge Argument targets physicalism about the mind, which claims that, as Jackson puts it in a 1986 follow-up article, "the actual world … is entirely physical". The argument provided one of the chief sources of doubt about physicalism in the late twentieth century, and continues to shape discussion of the mind-body problem into the twenty-first. It is unclear whether the argument converted many to dualism; still, most readers found the argument's core thought experiment highly compelling. Physicalists thus faced the challenge of identifying an error in the argument. The potency of the Knowledge Argument is clear from the fact that, while all materialists reject its conclusion, there is little agreement among them as to how, precisely, its reasoning is flawed.

The Argument

Jackson's original argument is disarmingly brief. He invites the reader to imagine the following scenario: Mary, a brilliant neuroscientist, has spent her entire life in a room in which the only visible colors are black and white. Partly through the use of a black-and-white television monitor, Mary comes to know all of the physical facts about color vision. These facts include the nature of causal interactions between the surface reflectance properties of objects, wavelengths of light, and retinal stimulation. Jackson then asks: "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?" He answers: "It seems just obvious that she will learn something about the world and our visual experience of it." He thinks that when Mary finally leaves the room and, for the first time, gazes upon an object that is red (and that she knows to be red), she learns what it's like to see red. Jackson concludes that, since physicalism requires that all facts are physical facts, physicalism is false.

Jackson's conclusion is a dualism of properties, rather than of substances; and this is all that his argument warrants. For a difference in properties—between the property instantiating neurophysiological state N, and the property instantiating qualitative state Q, say—suffices for a difference in corresponding facts.

A formalization of the argument will be useful.

1. While in the black-and-white room, Mary knows all of the physical facts about color experience.
2. Mary learns something about color experience upon her release.
3. If Mary learns something about color experience upon her release, she does not know all of the facts about color experience while in the room.
4. Mary does not know all of the facts about color experience while in the room. (from 2 and 3)
5. There are facts about color experience that are not physical facts. (from 1 and 4)
6. If physicalism is true, then all facts are physical facts.

Therefore,

7. Physicalism is false. (from 5 and 6)

As mentioned above, Jackson distinguishes this argument from Nagel's 1974 argument. Nagel had argued that no amount of physical information about bats—including knowledge of their neurophysiological, behavioral, and evolutionary features—could allow us to grasp the experiential aspect of using echolocation, that is, to know what it's like to be a bat. According to Jackson, these arguments differ in two ways. First, he claims that his argument concerns knowledge of a general property of experience, what it's like to see red, while Nagel's argument concerns knowledge of a property specific to an individual, that is, what it's like to be a (particular) bat. But to some, this difference has seemed at most a quirk of exposition: for Nagel's argument does draw into question whether we can know a general property of experience, namely, what it's like to use echolocation. However, others—including Jackson himself—have claimed that while Jackson's argument specifically targets the contrast between the phenomenal and the physical, Nagel's argument instead targets the contrast between the subjective and the objective.

The second point of contrast that Jackson draws is this: Nagel's argument simply shows that humans cannot imagine what it's like to use echolocation, and this limit to our imaginative powers is irrelevant to the issue of physicalism. Whether Nagel's argument rests on this issue about imaginability, or whether it would remain intact when using an experience that is within the normal course of human experience (as Jackson's does), is largely a question of interpretation. But the point about imaginability brings out an important and sometimes overlooked feature of Jackson's argument: that nothing in the argument excludes the possibility that Mary, perhaps through an exercise of imagination or as the result of taking a hallucinogen, undergoes an experience while in the room which is, in fact, a seeing red experience. Jackson's point remains so long as Mary is unable to determine that the experience is a seeing red experience as opposed to, say, a seeing green experience. This brings out the epistemic character of the argument. Jackson's argument requires only that Mary cannot deduce that a certain experience is the sort of experience her subjects undergo when seeing a ripe tomato (say). Upon leaving the room, Mary has the opportunity to correlate these, by gazing at a tomato herself. (She could, of course, correlate them while inside the room, by scanning her own brain while she is undergoing the seeing red experience. In the context of the argument, having the opportunity to make this correlation is tantamount to leaving the room.)

Objections to the Argument

I will focus on the four most influential types of objection to the argument. The first is simply to deny the conjunction of premises (1) and (2). On this view, Mary does not know all of the physical facts unless she knows what it's like to see red. Daniel Dennett (1991) takes this approach, arguing that we can't truly conceive knowing all of the relevant physical facts. This limitation explains why it seems that Mary learns something upon her release; but, Dennett maintains, if (1) is true, then (2) is false. In response, defenders of the knowledge argument have pointed out that the argument requires only that we understand the basic kind of knowledge that Mary has while in the room, not that we can mentally rehearse each bit of information she possesses. Since we do have a grasp of the sort of physical facts she knows, our powers of conceiving are strong enough to evaluate the possibility that (1) and (2) are true simultaneously.

The next two objections deny premise (3). Churchland describes what Mary gains upon her release as a kind of knowledge by acquaintance of what it's like to see red; that is, as non-propositional knowledge of this fact. Using this analysis of the case, he argues that a parallel argument would condemn dualism as well, since Mary would lack such knowledge by acquaintance even if she had exhaustive propositional knowledge about the nonphysical. Jackson (1986) responds that these are not on a par, for one could know all of the physical facts about seeing red without knowing what it's like, but one could not know all of the facts (physical and nonphysical) about seeing red without knowing what it's like. This may seem question-begging, but it has seemed highly intuitive to many philosophers, and hence this second avenue of objection has attracted relatively few proponents. (But see Earl Conee 1994 for a more developed version of the acquaintance analysis.)

Another objection that denies premise (3) claims that what Mary gains upon leaving the room is an ability, rather than knowledge of a fact. This objection originated in Laurence Nemirow's review of Nagel's argument (1980), and is defended by David Lewis (1988). On this ability approach to defusing the argument, when she finally sees something red, Mary learns how to remember, recognize, and/or imagine a seeing red experience. The fact that experience is required for such abilities carries no anti-physicalist consequences; after all, exhaustive propositional knowledge doesn't generally guarantee that one possesses the relevant ability. If it did, professional baseball teams would be staffed by physicists, who can master all of the relevant facts about how to hit a curve ball.

While the ability approach remains influential, it does face difficult challenges. One challenge is to specify an ability that is gained when, and only when, Mary learns what it's like to see red. At the moment of her grasping this, she is not yet able to remember what it's like, for the moment hasn't passed; and if she has a poor imagination, experience may not enable her to imagine what it's like. (For responses along these lines, see Conee 1994 and Torin Alter 1998.) Arguably, the best candidate for what Mary gains is the ability to recognize seeing red experiences. But the ability analysis may be mistaken even if this recognitional ability is perfectly correlated with knowing what it's like to see red. For, as Brie Gertler (1999) argues, it seems plausible that Mary is able to recognize a seeing red experiences *because* she knows what it's like, where because is

used in an explanatory sense. If knowing what it's like explains the recognitional ability, then it doesn't reduce to that ability.

The fourth and most widely accepted type of objection to the Knowledge Argument rejects premise (6). It claims that our ways of representing reality may be more fine-grained than the reality we represent, and what Mary gains is simply a new way to represent a portion of reality that was already known to her. (There are two competing ways to use fact in this context. One is to read fact as inheriting the fineness of grain that our representations possess; it is this reading I have used in saying that this objection targets premise (6). The second reading uses fact as less fine-grained than our representations. On that reading, the current objection would instead reject premise (3), claiming that Mary didn't learn any new facts but only encountered old facts under a new guise or mode of presentation. The difference here is purely verbal, and I will continue to use fact in the former sense.)

This sort of objection was present in earlier papers (including Terry Horgan 1984 and Michael Tye 1986), but is usually associated with Brian Loar, who provided a nuanced version of it in 1990. Loar argues that a single property may be the referent of distinct concepts. In particular, a property that Mary knew as instantiating neurophysiological state N may be identical to the property instantiating qualitative state Q, even if knowledge that a state falls under the former concept does not generate knowledge that it falls under the latter. Thus, Mary's ignorance can be attributed to a distinction in concepts that does not imply any distinction in properties.

More generally, this line of response to the Knowledge Argument construes the change in Mary as purely epistemic, and denies that her epistemic advance, upon leaving the room, reflects any grasp of a hitherto unknown ontological feature of the world. As such, it represents a more general, highly influential position about the mind-body problem: The apparent disparity between physical and phenomenal features of the world (called the explanatory gap after Joseph Levine (1983)) is purely epistemic, and not ontological.

This position belongs to a more general outlook known as a posteriori physicalism. According to a posteriori physicalists, anti-physicalist arguments that are based on thought experiments show, at most, that physicalism is not an a priori truth; but as Saul Kripke (1980) demonstrated, some identities are a posteriori (yet necessarily true). Strikingly, Kripke himself rejects a posteriori physicalism, and claims that the distinctive way in which phenomenal concepts operate rules out the possibility of a posteriori identities between phenomenal and physical (or functional) properties. In a paper co-written with David Chalmers, Jackson also objects to a posteriori physicalism. According to Chalmers and Jackson (2001), the approach used by a posteriori physicalists presumes that there is a deep schism between concepts and ontology, a schism that would undercut the justification for uncontroversial identity statements.

Despite his continuing opposition to a posteriori physicalism, Jackson now rejects the Knowledge Argument. He contends that phenomenal knowledge is deducible, in principle, from physical knowledge, even if we may be unable to perform the deduction. Jackson's turnabout is based on his acceptance of representationalism, which claims that the phenomenal character of a state is exhausted by its representational content. For instance, suppose that one of Mary's subjects, Joe, gazes at a ripe tomato. Representationalists maintain that the visual phenomenal

quality of Joe's experience is fully captured by the fact that his state represents there is something round and red before me. (Specific representational contents will be much more detailed, of course.) Since Mary can, in principle, know the representational contents of Joe's states before her release, she can in principle know all that there is to know about what it's like to see red.

Overall Assessment

The Knowledge Argument is an argument against physicalism. Yet its importance stems as much from the richness and variety of the responses inspired by its provocative reasoning as from its anti-physicalist conclusion. Discussion of the argument has profoundly affected debate on a range of issues, including: differences between propositional knowledge and ability, the relation between identity and deducibility, and the special features of phenomenal knowledge. While the majority of philosophers ultimately reject the argument, a vocal minority accepts it as sound.

Bibliography

Alter, Torin. "A Limited Defense of the Knowledge Argument." *Philosophical Studies* 90(1998):35-56.

Broad, C.D. *Mind and its Place in Nature*. London: Routledge and Kegan Paul, 1925.

Chalmers, David, and Jackson, Frank. "Conceptual Analysis and Reductive Explanation." *The Philosophical Review* 110(2001):315-360.

Churchland, Paul. "Reduction, Qualia, and the Direct Introspection of Brain States." *Journal of Philosophy* 82(1985):8-28.

Conee, Earl. "Phenomenal Knowledge." *Australasian Journal of Philosophy* 72(1994):136-150.

Dennett, Daniel. *Consciousness Explained*. Boston: Little Brown and Company, 1991.

Gertler, Brie. "A Defense of the Knowledge Argument." *Philosophical Studies* 93(1999)317-336.

Horgan, Terry. "Jackson on Physical Information and Qualia." *Philosophical Quarterly* 34(1984):147-52.

Jackson, Frank. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32(1982):127-136.

Jackson, Frank. "What Mary Didn't Know." *The Journal of Philosophy* 83(1986):291-295.

Jackson, Frank. "Mind and Illusion." In *Minds and Persons*, ed. Anthony O'Hear. Cambridge: Cambridge University Press, 2003.

Kripke, Saul. "Naming and Necessity." Cambridge: Harvard University Press, 1980.

Levine, Joseph. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64(1983): 354-361.

Lewis, David. "What Experience Teaches." In *Proceedings of the Russellian Society* 13(1988):29-57. Reprinted in Ludlow et. al, 2004.

Loar, Brian. “Phenomenal States.” In *Philosophical Perspectives IV: Action Theory and Philosophy of Mind*, ed. James Tomberlin. Atascadero, CA: Ridgeview Press, 1990.

Ludlow, Peter; Nagasawa, Yujin; and Stoljar, Daniel, eds. *There’s Something About Mary*. Cambridge, MA: MIT Press, 2004.

Nagel, Thomas. “What is it Like to be a Bat?” *Philosophical Review* 83(1974):435-50. Reprinted in his *Mortal Questions*. Cambridge: Cambridge University Press,.

Nemirow, Laurence. Review of *Mortal Questions* Thomas Nagel. *Philosophical Review* 89(1980):473-477.

Robinson, Harold. *Matter and Sense*. Cambridge: Cambridge University Press, 1982.

Tye, Michael. “The Subjective Qualities of Experience.” *Mind* 95(1986):1-17.