

Property Dualism and the Knowledge Argument: Are Qualia Really a Problem for Physicalism?

*Ronald Planer
Rutgers University*

Abstract: Where does the mind fit into the physical world? Not surprisingly, philosophers have offered radically different answers to this question. This paper considers and defends an argument to the effect that our conscious experiences must be something separate from the physical world since one could in principle know all the relevant physical and neuroscientific details without knowing anything about what it is like to actually have such experiences. Accordingly, I first present the premises and conclusion of the argument and then consider two popular ways of responding to it, highlighting some serious problems each of these strategies faces.

1. Introduction

The knowledge argument seeks to show that there are properties of our conscious experiences which are irreducibly mental. Roughly, it does this by showing that a complete physical description of everything which goes on in us when we have a visual experience – such as seeing the color red – is bound to leave out the qualitative nature of that experience; hence, a complete physical description must be incomplete. In this paper, I present the argument and consider two replies to it, namely, the Ability reply, and the New Knowledge/Old Fact reply. I argue that only the latter of these holds any promise for refuting the knowledge argument, but that even it looks worrisome in light of some of the criticisms I develop.

2. Consciousness

Consciousness is the thing we regain in the morning after a long, dreamless sleep. Or, alternatively, it is the thing we lose when we are “put under” before surgery.

In *The California Undergraduate Philosophy Review*, vol. 1, pp. 69-81. Fresno, CA: California State University, Fresno.

To be sure, there are many interesting aspects of consciousness, but if we hope to one day give an answer to the mind-body problem, we must be careful to distinguish them. In the philosophy of mind, debate is centered around what it called phenomenal consciousness or the phenomenal character of conscious experience.

Roughly, phenomenal consciousness is the what-it-is-like-for-a-subject aspect of consciousness. For example, there is something that it is like for us to hear our favorite song on the radio just as there is something that it is like for us to eat a good steak. Some of our experiences can be physically or emotionally painful while others bring us great joy. The point is that our conscious experiences have a certain intrinsic, qualitative character to them, and it is this aspect of consciousness that concerns us here. (Having made this qualification, we can drop the modifier "phenomenal" and simply speak of entities and states as being conscious. Hence, an entity is conscious if and only if there is something that it is like to be that entity, and a state is conscious if and only if there is something that it is like to be in that state.)

Our aporia over consciousness, then, is this: How could such "raw feels" in all their richness be reduced to something purely physical? How could our itches, tickles, and pains, be identical to some brain process or other? As hard as it might seem to wrap our heads around, physicalists nonetheless claim that the mental is nothing over and above the physical. Jackson, on the other hand, disagrees.

3. The Threat to Physicalism

In "Epiphenomenal Qualia," Frank Jackson mounts his attack on physicalism in two steps. First, by way of the knowledge argument, he establishes that a complete list of all the physical facts pertaining to our color perceiving apparatus and the nature of light is bound to be an incomplete description of what goes on in us when we actually perceive color. He then uses this conclusion to argue that physicalism is false. Let's consider each of these steps in turn.

3.1 The Knowledge Argument

Jackson asks us to imagine the following scenario: Mary is a super color perception scientist. That is, she knows all the neuroscientific facts pertaining to color perception in human beings as well as all the physical facts about electromagnetism and the color spectrum. However, ironically enough, Mary has never seen color

herself; she has been raised in a completely colorless environment, exposed to only black, white, and shades of gray.

But now suppose Mary leaves the room. She walks outside and for the first time sees a red rose. Jackson says that Mary learns something new; she learns what it is like to see red. But if that's true, then physicalism is false, since we assumed that before she left the room, Mary knew everything physical there was to know about light and our color apparatus.

More precisely, then, the argument goes:

(P1) Before Mary leaves the room, she knows all the physical facts about human color perception.

(P2) After leaving the room, she acquires new factual information about human color perception: namely, she learns what it is like to see red.

(C1) Therefore, it's not the case that Mary knew all the facts about human color perception before she left the room. [from (P1) and (P2)]

(C2) Therefore, it's not the case that all facts are physical facts. [from (P1) and (C1)]

This is the knowledge argument.

3.2 The Anti-Physicalist Argument

Jackson then goes on to use the ultimate conclusion of the knowledge argument, i.e., (C2), to argue for the falsity of physicalism:

(P3) If physicalism is true, then all facts are physical facts.

(C2) It's not the case that all facts are physical facts.

(C3) Therefore, physicalism is false. [from (P3) and (C2)]

This argument is obviously valid; however, one might object against the first premise, (P3). That is, why should physicalism entail that all facts are physical facts, facts about physical bodies, physical properties, and physical relations? Isn't this too strict a requirement?

Some philosophers, e.g., those that hold a non-reductive physicalist view of the mind, might think so. For example, perhaps they'd claim that all physicalism requires is that every *particular* mental state be identical to some *particular* physical state, and in this way allow that not all mental properties must be reducible to physical properties. At any rate, how to formulate the thesis of physicalism is itself a substantive philosophical issue that we cannot hope to shed light on here. We can rest assured that Jackson would be content to have shown that there are facts about our conscious experiences which a completely neuroscientific description is bound

to leave out, however one decides to define physicalism. Thus, (P3) need not detain us.

3.3 Before Criticizing Jackson . . .

To sum up: First, we asked whether consciousness, that is, the what-it-is-like aspect of our experience, could be physical. Then, in an attempt to answer this question negatively, we considered an argument to the effect that a complete physical description of what goes on in us when we see color is bound to be incomplete. This fact, together with (P3), led us to reject the thesis of physicalism in favor of some form of property dualism, i.e., that there are non-physical properties to our conscious experiences.

Jackson thinks this line of reasoning is sound. However, are there really no problems with it? Two questions rise to the surface in assessing the strength of his argument:

(Question 1) Surely Mary has a new experience when she leaves the room, but does she really acquire any new factual information? and (Question 2) Supposing she does acquire new factual information, is this really a problem for physicalism?

The strategies for responding to Jackson are conveniently framed as answers to these questions, i.e., the Ability reply answers Question 1 negatively, and the New Knowledge/Old Fact reply answers Question 2 negatively. Let's see how they go.

4. The Ability Reply

In his (1988), David Lewis responded to the knowledge argument as follows. Boiled down, his idea is to distinguish between knowledge-that and knowledge-how. *Knowledge-that* is knowledge that something is the case, i.e., it is the kind of knowledge we have when we grasp the significance or import of a proposition such as, $2 + 2 = 4$. In contrast, *knowledge-how* is a matter of having certain abilities. For example, when we know how to drive a car or play poker, we possess the ability to do something. But knowing how to do such things clearly does not amount to factual knowledge. (If I learn how to ride a bike, I do not acquire any new factual information.)

With this distinction in place, Lewis then says that after leaving the room, Mary acquires new knowledge-how, she acquires certain new abilities she didn't have before she left the room, but she does not acquire any new knowledge-that. What abilities does she gain? Once Mary leaves, she gains the ability to imagine the color red to

herself, to recognize future instances of red, to compare red with black and white, and so on. As such, Lewis's strategy amounts to a denial of (P2), i.e., the premise which states that Mary acquires new factual information (or knowledge-that) when she sees a red rose. This allows us to hold on to the idea that before leaving the room Mary's knowledge of what goes on in us when we perceive color is complete. Hence, physicalism is safe.

Counter-reply: The Ability reply misses the point for the following reason. It's certainly true that Mary acquires new abilities when she leaves the room, and thus knowledge-how, but she also acquires new knowledge-that, specifically, the knowledge that *that* is what it is like to see red or to see red is like *that*.

What this counter-reply states is that knowledge of what it is like to see red cannot be analyzed solely in terms of the ability to imagine red to oneself or to recognize future instances of red. This is because, intuitively, we may completely lack these abilities and still know what it is like to see red (Conee, 1994; Alter, 1998). For example, suppose Tom has suffered a head injury such that he is completely devoid of the ability to imagine red to himself. Suppose also that even though Tom has seen red many times, he cannot pick out future instances of red from other colors due to this damage. Despite Tom's handicap, it seems implausible to say that when Tom is actually seeing a red patch, that is, when we hold a red patch directly before his open eyes, and we tell Tom that what he is presently seeing is red, that he does not know what it is like to see red. Therefore, what it is like to see red can't just be a matter of having certain abilities. Rather, it is also involves the factual knowledge that *that* is what it is like to see red.

But even if one is unconvinced by the above thought-experiment, I think there is a more serious problem with the Ability reply that can be brought out as follows. First, even though knowledge-that and knowledge-how are clearly distinct from each other, they are often intertwined and related in complicated ways. For example, consider the process of reading and understanding a sentence. This task involves both knowledge-that, i.e., it involves knowledge of certain words, grammatical rules, social conventions and other expectations on the part of the reader, but it also obviously involves knowledge-how as well, i.e., it involves the ability to visually perceive the markings on the page as letters, to process that information so that it can then be made available for us to interpret and understand in a meaningful way. Thus, reading is a complex act in which both kinds of knowledge are made use of.

What does this have to do with the Ability reply, though? Similarly, it seems that the abilities Mary is supposed to have gained as a result

of leaving the room are also knowledge-that/-how compounds. For example, consider the case of the ability to recognize future instances of red. What we do when we actually recognize something, be it the color red, our house, or our own face in the mirror, is to judge whether or not it is familiar to us, that is, whether we have encountered it before, and this is done on the basis of the thing perceived sharing enough of its properties in common with a mental representation we have stored in our memory of that thing. Thus, even if we are not consciously aware of this decision-making process (obviously it does not seem to us that we are judging whether or not that is our face in the mirror!), it surely does go on in us, as evidenced by the mistakes we sometimes make with respect to recognizing familiar things. The problem for the Ability reply, then, is that the information with which we make such judgments (perhaps "responses" is a better word here) is almost certainly knowledge-that, e.g., recognizing that building over there as my house involves the factual knowledge that my house is two-stories tall, is painted blue, is located between the Smiths' and the Jones' houses, etc. In short, it seems implausible that Mary could have the ability to recognize future instances of red without also having the factual knowledge that to see red looks like *that*. And as for *imagining* instances of red to oneself: it seems that it is precisely the fact that you know what red looks like, that is, that among the many propositions you know, you know red looks like *that*, that you are able to conjure the appropriate image up before your mind. Thus, what this objection shows is that at least as often as knowledge-that and knowledge-how are clearly separable from one another, they are also often connected in complex ways, the one making the other possible, and vice versa, and this seems to be the case especially with complicated cognitive tasks such as recognition and imagination, in contrast to mere brute skills, such as walking or chewing your food.

5. The New Knowledge/Old Fact Reply

One version of New Knowledge/Old Fact reply is taken by Brian Loar (1990). The advantage of this strategy over Lewis's is that it can accept the intuition that Mary acquires new factual knowledge when she leaves the room. However, Loar claims that this does not imply there are non-physical facts.

How might this work? That is, supposing that Mary knows all the physical facts about human color perception before she leaves the room, how would there be any room left over for information she is ignorant of but which is nonetheless physical? The problem with the knowledge argument, according to Loar, is that it collapses the

distinction between learning new factual information and learning a new fact. This point is subtle and best brought out by analogy with simpler cases. For example, suppose Sally knows that water is wet. Now suppose she takes a chemistry class and learns H_2O is wet. Presumably, Sally does not come to know a new fact. Rather, the knowledge she gains is knowledge of an old fact (i.e., that *water* is wet) seen in a new way. Or alternatively, suppose Bill knows that the morning star is hot. Suppose further that he then finds out that the evening star is hot. Bill certainly possesses new factual knowledge, knowledge-that, but it is knowledge he already had under a different description (i.e., that *the morning star* is hot).

This allows us to give the following analysis of Mary's epistemic situation. When Mary leaves the room, she acquires new factual information, namely, that *that* is what it is like to see red or to see red is like *that*. But this is merely knowledge of an old fact seen in a new way: i.e., whereas before she left the room, Mary knew what it is like to see red under an objective, neuroscientific description, say, the occurrence of neuronal process X4978, when she actually sees red first-hand, she learns the same fact but under a different description, a demonstrative one. Hence, although Mary acquires new factual information when she leaves the room, this does not imply that she learns a new fact. Therefore, we are not forced to give up the assumption that Mary possessed all the *facts* about human color perception while still in the room. (This amounts to saying the inference from (P1) and (P2) to (C1) invalid.)

Counter-reply: The New Knowledge/Old Fact reply presupposes the idea that we can know a fact F under one description or mode of presentation, D_1 , but not under another description or mode of presentation, D_2 . More precisely, in order to defeat the knowledge argument, the descriptions "neuronal process X4978" and "*that*" (where "*that*" demonstratively refers to the immediate phenomenological character of Mary's red-seeing experience) must serve to pick out one and the same fact, namely, what it is like to see red.

The problem with this is as follows: The identity statement, "neuronal process X4978 = *that*" is clearly an empirical one, one Mary had to "discover." That's to say, no amount of *a priori* reasoning could ever have told her that these terms were co-extensive. What explains this epistemic gap? The normal story to tell is that there are two logically independent sets of properties, namely, those properties which guide Mary in picking out what it is like to see red under its phenomenological mode of presentation, and those properties which guide Mary in picking out what it is like to see red under its objective, neuroscientific mode of presentation, and it just so

happens that these two sets of properties pick out one and the same thing. After all, this seems like a plausible explanation for why an identity statement such as "the Morning Star = the Evening Star" had to be discovered. But this explanation isn't open to proponents of the New Knowledge/Old Fact reply on pains of conceding that there are mentally irreducible properties, or, in other words, that property dualism is true. This is because in order to hold on to the idea that all the properties of our conscious experiences are physical properties, they must also defend the claim that the *property* of being neuronal process X4978 is identical to the *property* of being *that*. Thus, the challenge to those who take this line of response to the knowledge argument is that they must provide us with an alternative explanation to the one sketched in this paragraph as to why the identity statement "neuronal process X4978 = *that*" is an *a posteriori* one. Can this be done?

There are two ways one might respond to this objection. The first is to claim that our talk of conscious mental states is topic-neutral. The notion of using such "topic-neutral analysis" to handle the problem of qualia comes from J.J.C. Smart (1959), in response to an objection famously put to him by Max Black. The second is a more recent strategy, one that involves distinguishing concepts from properties.¹ Let us consider each of these alternatives in turn.

At the heart of the topic-neutral approach is the idea that when we report we are experiencing states of phenomenal consciousness, we report merely that there is something going on in us which . . . (where ". . ." gets filled in by a list of functional properties, depending on the mental state we are reporting.) So, for example, consider the report that I am in pain. When I think or say aloud the sentence "I am in pain," what I am really reporting is that *there is something going on in me which is a sign of physical damage being done to my body, which is apt to cause avoidance-behavior in me, which is apt to make me shout "Ouch!"* and so on. And as the word "something" merely serves as a place-holder in this description, this formulation of what it is for a state or process to be the feel of pain obviously leaves open the question as to whether the nature of the "something" I am reporting is physical or non-physical (hence the name "topic-neutral").

1 This strategy was first suggested to me by Professor Brian McLaughlin of Rutgers University. We were discussing the challenge to the identity theory implicit in Nagel's "What Is It Like to Be a Bat," (1974). Roughly, the challenge is this: (1) Mental states are essentially subjective; (2) Brain states are essentially objective; (3) How can something be both essentially subjective and essentially objective? Of course the thing to say here is that a state or process is only subjective/objective under a certain concept of that state. For more on this strategy, see Loar (1990).

Hence, in short, the topic-neutral approach says that our conception of mental states is really very abstract, that "talk of the mental" just is talk of "something" which goes on in us under conditions *C* and is apt to cause effects *E*.

How does this solve the problem sketched in our counter-reply? That is, how does this strategy account for the fact Mary had to discover via empirical means that "neuronal process X4978 = *that*"? Roughly, the idea is this. *Qua* its phenomenological mode of presentation, what it is like to see red is presented only very abstractly to us, as something which goes on in us when we see a red rose, etc. However, in contrast, the same phenomenon *qua* its objective, neuroscientific mode of presentation, is not presented abstractly to us at all, but rather as a specific neuronal process that takes place in the brain. Thus, were we to find out that the identity statement "neuronal process X4978 = *that*" is true, it would simply be a case of discovering which brain process is picked out by the "something" in the mental description of the event. And since this is a discovery that can only be made by empirical means, it would explain how it is that Mary can know all about the neurological event without knowing that to see red is like *that*. However, are there really no problems with this approach?

Although the topic-neutral reply does provide us with one explanation of the epistemic gap, I think it is ultimately inadequate for the following reasons. First, what appears to be the only essential property of many mental states in general and states of phenomenal consciousness in particular is that inner, subjective, qualitative feel of the state, and Smart's treatment of the mental seems to ignore this aspect our experiences altogether, substituting for it some functional role that the state plays instead. If I am right, then we must worry that a state could meet all of the appropriate functional conditions for a certain mental state without having the accompanying what-it-is-like-for-a-subject aspect to it. The second problem is that we are obviously not guided in picking out our mental states, say, the feel of pain, by any such topic-neutral description. Rather, we focus only on the very feel of the state itself – *anything that feels like pain is pain, and nothing that does not feel like pain is pain, end of story*. Particularly this last point I think shows that our conception of mental states cannot be as abstract as Smart claims they are. This suggests to me that even if the topic-neutral approach is a consistent reply to our objection, it is ultimately bound to fail on independent grounds. So much for this alternative, then. What about the other?

The second line of response to our counter-reply is to distinguish between concepts and properties. Roughly, a concept is a way we have of thinking about or conceiving of a particular thing as being.

In contrast to this, a property is way a particular thing might actually be independent of our ideas about it. To illustrate this distinction, then, consider water. Presumably, the property of being water and the property of being H_2O is the same property – any possible world in which a substance X instantiates the property of being water, it also instantiates the property of being H_2O , and vice versa. Yet, it is not hard to see that our concept of water is very distinct from our concept of H_2O : i.e., we think of water as being a clear, odorless liquid, as something we drink, wash dishes with, and bathe with, whereas we think of H_2O as situated within a theoretical framework, that is, as a chemical compound consisting of one molecule Hydrogen and two molecules Oxygen. But as we all know, as distinct as these concepts may be, they pick out the same property in nature, namely, that of being water/ H_2O . So, instances of water are identical with instances of H_2O , the property of being water is identical to the property of being H_2O , but the concept of water is distinct from the concept of H_2O .

So, in response to our worry for the New Knowledge/Old Fact reply, we can say that while Mary's concept of *that* (i.e., her phenomenological concept of what it is like to see red) and her concept of neuronal process X4978 pick out the same property, they are themselves very distinct, even more distinct than our concepts of water and H_2O . This would explain how it is that Mary could know that to see red is to have neuronal process X4978, without knowing that to see red is like *that*; she has two radically different ways of conceiving of one and the same property, one which involves an objective, neuroscientific way of thinking, the other which involves a subjective, demonstrative way of thinking. It just so happens that the reference-fixing properties which guide Mary in picking out the referent under each mode of presentation pick out one and the same property, namely, what it is like to see red. (One might worry at this point whether or not the concepts/properties approach has lapsed into some form of property dualism. That is, by conceding that there are two independent sets of reference-fixing properties corresponding to her two concepts, haven't we also conceded that there are both mental and physical properties? This threat is more apparent than real, though. This is because the *only* property which guides Mary in picking out what it is like to see red under its phenomenological mode of presentation is the property of being *that* very look, and we're assuming for the sake of argument that this property is identical to the property of being neuronal process X4978.)

Thus, this approach too offers us a clear explanation as to why such property mental-physical identity statements are expected to be

empirical. Also, it fares well with respect to our criticisms of the topic-neutral approach: whereas the topic-neutral approach seemed to have a problem both with the what-it-is-like aspect to our conscious experiences and with the fact that we do not use any such abstract description in picking out our mental states, this approach doesn't. This is because it makes no claim that concepts of phenomenal consciousness are conceptually analyzable or reducible to physical/functional properties. This is really the key move here. As such, it does not conflict with the claim that mental states such as pain, are intrinsically subjective and qualitative, nor does it conflict with the claim that the only property with which we pick out pain is the very feel of the pain itself. It's just that a state or process has these features only under a particular phenomenological concept of that state/process. Accordingly, we cannot infer from the observation that the property identity statement "neuronal process X4789 = *that*" is false simply on the basis that there is no way to know this independent of experience, for it may simply be a case of conceiving of one and the same property under two different concepts, like water and H₂O. But are there no other problems facing this approach?

One concern that rises to the surface at this point is this. We use concepts to group together diverse instances we observe at different times. In some cases, the grounds for grouping these instances together are functional, i.e., we discover that each time we perform a certain action towards such-and-such an object, the same consequences are apt to result, while in other cases it is in virtue of the fact that instances share certain perceptual features in common that we group them together, i.e., owing to some feature of the object or property itself, and owing to the way our nervous system gathers and analyzes that information, diverse instances appear similar and come to be associated with one another.

Now the problem is that there seems to be an important difference between pairs of concepts such as "water" and "H₂O" or "lighting" and "electrical discharge" and phenomenological concepts such as "the feel of pain" or "the sight of a red rose" and concepts of brain states and processes: namely, the laws of physics and chemistry in conjunction with facts about the way our perceptual system automatically organizes sensory information predict and explain why a bunch of H₂O molecules will be perceived by us as a clear, odorless liquid, and, additionally, the laws of physics in conjunction with certain neurophysiological facts also predict and explain why under the appropriate conditions electrical discharge will be perceived by us as a bright flash of white light; however, it's very hard to see how such natural and perceptual laws could ever

predict or explain why it is that the sight of a red rose should have the inner, qualitative character that it does, the raw feel of the state that we all know so well. That's to say, a complete body of physical laws together with a description of our perceptual system would presumably vindicate our concepts of things such as water and lightning by showing why the reference fixing properties of those concepts are to be expected, yet in the case of neuronal process X4978 and Mary's phenomenological concept *that* it seems we lack the necessary schemas to even imagine how such an explanation might be given by the empirical sciences. Thus, it's not clear that the intuitive force which the concepts/properties strategy derives from its ability to account for the empirical nature of property identity statements such as "water = H₂O" or "lightning = electrical discharge" should transfer over to its ability account for property identity statements between phenomenological terms such as "*that*" and brain terms.

6. Conclusion

In this paper, I presented the knowledge argument and considered two replies to it, namely, the Ability reply and the New Knowledge/Old Fact reply. In light of my criticism of both replies, it seems to me that only the latter of the two has any chance at defending physicalism in the face of the qualia which the knowledge argument turns on.

Within our discussion of the New Knowledge/Old Fact reply, we objected that even if what Mary learns upon leaving the room is not a new fact, it still remains to be shown why she couldn't possibly have known beforehand that to see red would have phenomenological character that it does. What this problem really boils down to is how to account for the fact that property identity statements between phenomenological terms and brain terms are knowable only *a posteriori*, and any adequate reply to the knowledge argument I think must answer this question. The problem for the physicalist is that there seem to be just so many difficulties along the way.

REFERENCES

- Alter, T. (1998). "Mary's New Perspective." *Australian Journal of Philosophy*, 72: 582-584.
- Conee, E. (1994). "Phenomenal Knowledge." *Australian Journal of Philosophy*, 72: 136-150.

- Jackson, F. (1982). "Epiphenomenal Qualia." In D. Chalmers *Philosophy of Mind: Contemporary and Classical Readings*, pp. 273-280. New York: Oxford University Press.
- Lewis, D. (1988). "What Experience Teaches." In D. Chalmers *Philosophy of Mind: Contemporary and Classical Readings*, pp. 281-294. New York: Oxford University Press.
- Loar, B. (1990). "Phenomenal States." In D. Chalmers *Philosophy of Mind: Contemporary and Classical Readings*, pp.295-311. New York: Oxford University Press.
- Smart, J. (1959). "Sensations and Brain Processes." In D. Chalmers *Philosophy of Mind: Contemporary and Classical Readings*, pp.60-68. New York: Oxford University Press.