

Alexander Janiak

10/31/2022

### Mary's Red Apple: A Critique of The Knowledge Argument Against Physicalism

Enter Mary – a scientist from a future where all physical facts have been discovered including all casual, relational, and functional roles of completed physics, chemistry, and neurophysiology. Mary lives her entire life within a black-and-white room and learns all the facts of physics by watching lectures in black-and-white. Her knowledge, consequently, includes all there is to know about human color vision and the color red. One day, she leaves the room and sees the color red for the first time. Does she learn something new in her experience of seeing red (Alter)?

This hypothetical was originally framed by Frank Jackson as an argument against physicalism – the idea that everything in the world is physical – and for dualism – the idea that is a non-physical component to the world in addition to the physical. Intuitively, it seems that in the hypothetical Mary learns something new: what it's like to see red. In other words, she learns the *qualia*, or the properties that characterize what it's like, of red. But if she learns something new from this experience, this would mean that even if you were to know everything physical about something, you wouldn't know *everything*. This seems to imply that there is something other than what is physical, proving physicalism false. In fact, this argument is called the knowledge argument against physicalism. This argument, while intuitively against physicalism, actually fails to disprove it. It can be refuted by claiming the situation is impossible – Mary can't actually learn all there is to know about the physical in her black-and-white room. Furthermore, even if the situation is possible, it shows that if you know everything physical about something,

experiencing it for the first time gives knowledge of how to recognize that thing or what that thing is like – rather than a new property of that thing.

There are multiple responses to this argument, some of which are strong enough in their rebuttal to successfully cast doubt on the original argument. The simplest of these rebuttals is to attack the assumption that Mary can learn everything there is to learn about the physical through a black-and-white television (McCain). This rebuttal makes use of Hempel's dilemma which suggests that no one has a clear understanding of what constitutes a physical property. That is, it is impossible to know what constitutes the rest of the physical world from our current incomplete model of the universe. You can't make use of our current model of physics as the model for physicalism because it is clearly not complete. You also can't make predictions about the future of physics because who is to say what constitutes the rest of physics (Stoljar)? Using the dilemma as leverage, it's easy to see how making a claim about whether Mary can learn all there is to know about the physical through black-and-white lectures makes a large assumption. In fact, it's difficult to even analyze whether Mary learns something new if one is unable to define a physical property. For example, if the future of physics contained mental properties, then Mary wouldn't learn anything new from seeing red for the first time. This therefore invalidates the knowledge argument against physicalism.

The rebuttal using Hempel's argument, while convincing, is unsatisfying. It suggests we are unable to find the answer to the knowledge argument until we know we have completed the entire model of physics. Furthermore, it ignores the idea that we might still have some concept of what is physical even if we don't have the full model of physics. You can still reasonably assume that the chair you sit on is made of matter and is physical. It might also be reasonable to assume

that our current understanding of colors and light is complete enough to keep questioning the argument.

A more compelling rebuttal stems from claiming that Mary gains non-propositional knowledge, i.e knowledge other than facts, from seeing the color red. First it's important to delve deeper into what might happen when Mary sees red after leaving the black-and-white room. Before Mary leaves the room, it can be assumed that she can't produce a mental image of the color red; that is, she doesn't know what red looks like because she has never seen it. When she looks at red, the photons traveling from the red object reach her retina triggering a chain of neurons which cause her to recognize red and see red within her visual field. But where is that visual field? When she creates a mental image of red, where is it? In the physical sense, it seems there is no single thing you can point to and claim that it is a visual field or mental image.

This is why intuitively, it seems as if there are two properties at play: the physical/electric encoding of the color red and Mary's non-physical experience of red. The chain of neuron's firing from seeing the red is the electric encoding of Mary's received visual information. The visual experience, on the other hand, is the experience encoded from the chain of neuron's firing. Under this model, Mary is learning about the non-physical properties associated with seeing red. She learns the experience of seeing red. It's evident that accepting that Mary learned something new implies dualism. To counter this claim, one must examine whether the experience Mary has is genuinely novel and whether it teaches her something new about the properties of the color red.

A response to the argument is that Mary, instead of learning something new about red, learns acquaintance knowledge – i.e, she becomes acquainted with the experience of red. This is not new knowledge that she can gain by studying physics and science of the color red (McCain).

Instead it is subjective to her experience of red. In fact, it is not part of the properties of the color red. For example, imagine a scenario where there is another scientist, John, that is placed in the exact same setting as Mary. He learns all there is to know about physics and lives in a black-and-white room his entire life. The difference is, when John steps out of the room and looks at the red object, his visual field sees the red object as being Mary's experience of blue. He still becomes acquainted with red, but to Mary, he sees a completely different color. Such an example shows that although both Mary and John are looking at a red object, they can both experience red differently. Thus, the experience of the color red cannot be a property of the color.

According to the externalist point-of-view, the experience gained from seeing the color red would be unjustified using just bare visual senses, and thus would not constitute new propositional-knowledge. Externalism suggests that justification depends on factors that are external to a person. Under this view point, by John experiencing the color blue, the experience can't constitute new propositional knowledge because he is actually looking at a red object. Therefore, that new propositional knowledge would be unjustified. Instead, John gains acquaintance knowledge relating to his experience of red.

Furthermore, it can also be argued that Mary instead gains knowledge-how – i.e., knowledge of how to recognize or imagine. Mary learns how to identify red but learns nothing new about the color red, the reason being because John and Mary would have completely different knowledge of how to identify red. John identifies it by seeing Mary's blue while Mary identifies it by seeing red. An example of such knowledge-how can more easily be understood by investigating the knowledge of being able to identify a table. You can learn all the physical properties of tables – i.e., more than one “leg” connected with the ground with a flat “table-top” covering it – and know all there is to know about tables. Once you see a table for the first time,

you don't learn anything new about tables. You just learn how to recognize them in a reliable subjective way. It doesn't matter if one person first counts the "legs" then investigates the "table-top" or whether another person makes sure the "legs" are connected to the ground first – both have the same knowledge of what a table is. Their methods of identifying a table are just reflections of what a table is; they don't define what the table is.

It's also important to note that the knowledge-how gained by John would be protected from externalism because it is still a reliable way for John to identify red. That is, assuming all colors are unique, John would still be justified in identifying red by seeing blue because everytime he sees blue he associates it with red. Thus, according to the externalist point-of-view, Mary would ultimately gain non-propositional knowledge not associated with any property of red.

But what about the internalist point-of-view? This case is a little trickier. If Mary is an internalist, she believes that knowledge is justified through faculties and reasons internal to her. When she first looks at the red object, she appears justified in believing that the experience of red is a facet of what red actually is. The experience of red seems to define what the color red is for Mary. Given this view, it can be said that if Mary is an internalist, the knowledge argument proves successful; Mary learns new non-physical propositional-knowledge about the color red. This argument, while tempting, doesn't include the full-breadth of Mary's knowledge after leaving the room. Mary knows that the color red can be defined as an electromagnetic wave traveling at a certain wavelength. She can therefore recognize that an object is red without having to experience it. When she experiences the color red then, it is her retina that recognizes the object red by the physical properties of red and then encodes it into the experience of red. She knows this is what happens when she sees red. Therefore she isn't learning something new

about red, because her experience is based on the physical properties of red. Thus she doesn't gain any new propositional-knowledge.

Like in the externalist case, Mary could be gaining acquaintance knowledge or knowledge-how from the internalist view point. The argument for knowledge-how is similar. Mary experiencing red for the first time gives her a new way to identify the color red using experience - created from her brain processing the physical properties of red. For acquaintance knowledge, Mary learns what it is like to experience red but no property of red that she didn't already know.

The previous line of reasoning shows that Mary, rather than learning something new about the color red, gains non-propositional knowledge from experiencing the color red for the first time through both the lens of the externalist and internalist. This approves the intuition that Mary does in fact learn something when she walks out of the room but does not necessarily enforce physicalism. It simply counters the knowledge argument against physicalism and prevents the acceptance of dualism. This line of reasoning, however, is not without its potential flaws.

A rather interesting counter could be to ask whether the experience of something can be integral to itself. For example, if someone were to learn every possible physical fact about you – map out every neuron and atom in your brain – would they truly know you without ever experiencing what it is like to be you? Intuitively it might be that you identify more with your experiences than any other part of you. If so, then experience can hold novel propositional knowledge that can be a defining property of that thing. In terms of Mary's example, the experience of seeing the color red could be a defining property of red. That is, experiencing a color could be a fundamental property of what defines that color. The main issue with this

counter is it assumes that experiencing something is equivalent to experiencing what it is like to be that something. The question should instead be, although nonsensical: what is it like *to be* the color red? In addition, it assumes that what is gained from experiences cannot be gained from knowing everything that is physical about something. Perhaps by learning all the neuron mappings of a human, you have access to all their experiences through their memories and neural connections.

A more direct attack could be made by bringing up another related argument: the conceivability argument. To prove that the experience of red isn't a property of red, a hypothetical scenario was conceived of John, someone in the same scenario as Mary, who sees red as Mary's blue. Such a hypothetical scenario would describe John as having inverted qualia compared to Mary. If this scenario is conceivable, it does not seem inconceivable that there could be a world exactly like ours but without experiences and consciousness. If such a world is conceivable, then there could exist creatures without experiences and consciousness, and thus proving physicalism false: consciousness is neither physical or functional (Alter). Such an argument is dangerous because it uses the same facilities of conceivability as the initial line of reasoning. That being said, it is much more conceivable to imagine two people with inverted qualia because there are real life examples of such scenarios. For example, color blind individuals see the qualia of colors differently than people without color blindness. On the other hand, although it is conceivable to imagine people without consciousness, it doesn't not necessarily mean it is possible.

The knowledge argument attempts to argue that even if you know everything physical about something, more can be learned through experiencing it. This implies that there must be another substance or property different from the physical which allows for additional

propositional knowledge. This argument can be refuted – unsatisfyingly – by casting doubt on whether we can assert that one can learn all there is to know about what is physical under certain constraints without knowing what the future of physics holds. Furthermore, it claims that one can't even make predictions about what constitutes a physical property without the complete model of physics. A more compelling argument is to claim that rather than learning propositional knowledge, Mary – who knows about everything physical – gains non-propositional knowledge regarding her experience of the color red. This non-propositional knowledge can either be acquaintance knowledge or knowledge-how. This can be potentially refuted by arguing that experience is an essential component of the color red or by utilizing the conceivability argument against physicalism.



Works Cited

Alter, Torin. "Knowledge Argument Against Physicalism." *Internet Encyclopedia of*

*Philosophy*, <https://iep.utm.edu/know-arg/>.

McCain, Kevin. "Mary Sees an Apple (Knowledge Argument Against Physicalism)."

*Epistemology: 50 Puzzles, Paradoxes, and Thought Experiments*, Routledge, New York, NY, 2022.

Stoljar, Daniel. "Physicalism." *Stanford Encyclopedia of Philosophy*, Stanford

University, 25 May 2021,

<https://plato.stanford.edu/entries/physicalism/#HempDile>.