

The Metaphysical Mind Flaw

My Physicalist Response to the Knowledge Argument

Ikshvak Tanay Ongolu

May 2026

Introduction

The mind-body problem is one where it is debated whether the mind and body are one, or whether there exists another metaphysical entity (the mind) which derives certain properties, or is in some way related to the body (a.k.a. the Brain). In this article, I will be examining Jackson's Knowledge Argument from *What Mary Didn't Know* (1986), and arguing in favor of my physicalist perspective on the mind-body problem. Physicalism is defined as the belief that the real world consists simply of the physical world. Essentially, that there exists nothing that isn't rooted in some fashion to physical, observable things. I believe that Jackson's knowledge argument fails, because it incorrectly assumes that Mary already contains, and will contain, every bit of physical information about red, and seeing red, before her release from the black-and-white state. I will first lay out Jackson's argument as he presents it, then target his first premise by showing that it contains an impossibility that isn't merely physical but logical, then address how Jackson would likely respond using his own framing of the argument, and finally close with a thought experiment that I think leaves nothing for Jackson's nonphysical idea of color to latch onto.

Jackson's Knowledge Argument

In Jackson's Knowledge Argument, he describes Mary as a woman that "... is confined to a black-and-white room, is educated through black-and-white books, and through lectures relayed on black-and-white television... she learns everything there is to know about the physical nature of the world ... [including] everything in completed physics, chemistry, and neurophysiology" (Jackson, 1986). This is essentially the setup of the thought experiment Jackson lays out, where Mary is described as essentially omniscient about color, except for the physical state when viewing it. He then states, "when [Mary] 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 ... hence, physicalism is false" (Jackson, 1986). The main argument in this case is that, if Mary already had all the physical facts and yet still learns something upon her release from this black and white room, there must be something nonphysical (a.k.a, physicalism is false). Jackson's knowledge argument essentially states:

Premise 1: Mary knows all the physical information that there possibly is to know about color experience before her release into the colored world.

Premise 2: When exposed to a colored object, Mary learns something new about the experience of color.

Premise 3 (implied): The new thing that Mary learned when seeing color was something other than physical information.

Conclusion: Therefore, Physicalism is false.

My Physicalist Rebuttal

My objection targets Premise 1. Jackson's argument is fundamentally flawed, as he rests the validity of physicalism on knowledge. He essentially argues that if one already has all the knowledge physically known of something, then experiences that same thing (she now "knows what it's like" to see the color red), then her experience of what red "is like to see" must not be physical. I will push back against this argument, because I would argue that the experience of seeing the flower is a fully physical experience.

There is absolutely no metaphysical component to seeing the flower and feeling what it is like to see it. The human experience is fundamentally rooted in physics. The emotions, feelings, experiences, and micro-moments associated with the red light hitting her retina are all physically rooted in complex arrays of neural wirings and feedback loops that propagate with other feedback loops which associate emotion, knowledge, and thought into a complex behavior we then call experience or feeling. So in spite of Mary having all the physical knowledge there ever was to know about the color red, she doesn't learn something metaphysical. She actually learns new physical information that isn't possibly teachable, as this is an involuntary physical response when confronted with the physical wavelengths of the color red. If Mary is to know all the physical information about red, then this must include the physiologically generated physical information when actually seeing red. Mary in the room has all the physical information that can be communicated through language, but not all the physical information.

To solidify my position, I am not only arguing from a current physical-possibility sense that Mary cannot learn everything about red without seeing it. Instead, I am simply pointing out a logical impossibility. It is logically impossible to know everything about seeing color without actually seeing color, the same way it is logically impossible to be a married bachelor. Jackson is simply contradicting himself by saying Mary knows everything there is to know about the color red, without her actually seeing red.

Jackson's Potential Response

Now, Jackson may likely respond that I am essentially, indirectly rejecting the hypothetical he has stated. The whole point of Jackson's argument is that Mary is defined as having all physical information, and if I say she doesn't, I'm just rejecting Jackson's hypothetical. Jackson claims that the argument "is concerned with the nature of Mary's total body of knowledge before she is released" (Jackson, 1986), and argues that Mary's whole database of physical knowledge of color was incomplete in my argument.

However, I am not rejecting Jackson's hypothetical. I am simply arguing that the hypothetical contains an inherent impossibility. It is impossible for Mary to have "all the physical information about color" without undergoing the physical process of seeing color, because that process is itself physical information. Let's say, for example, in the same scenario of Mary's black-and-white room, Mary is given all the knowledge of the actual experience of looking at the flower. As in, there is an expert neuroscientist who has somehow calibrated a device perfectly to Mary's brain, where activation of the device will light up the exact sequence of neurons, their wirings, and their associated cascades, identical to what would happen if Mary were to see the color red. The only missing stimulus would then be the red color wavelength actually hitting Mary's retina. In this case, I could still deny Jackson's argument for a metaphysical process of experience. Because if this new neuro-experienced Mary were to then be let out and see the color red, there would literally be nothing new she would have learned. Every part of her physical anatomy was calibrated precisely, and activated previously by the neuroscientist. If the experience of red can be fully produced by physical manipulation of the brain, there is nothing left over for Jackson's nonphysical "something" to be.

Now Jackson could still refute me, and say that I am incorporating the experience of seeing color itself rather than transmitted information about it. But I think if Jackson were to say that, he would have immediately proved my point. If activating the right physical states is the experience, then "information about the experience" and "the experience itself" can't be cleanly separated in the way Jackson needs them to be for P1 to even be coherent. The neuroscientist scenario is just making explicit that the physical facts about the experience and the physical realization of the experience are the same thing under different descriptions. There is no leftover non-physical phenomenon for Jackson's "something" to occupy.

Conclusion

Humans view the world almost through a window, where we know physical things by description. We experience them through our senses, and our emotions are simply reactions of neural cascades in complex, multifaceted manners, to elicit and alter certain behavioral (explicit or implicit) responses

within our brains. Our feeling when we see a red flower, for example, despite already being the foremost experts of the said red flower, is characterized simply by the manner in which our neurons fire in different regions of our brain reminiscent of emotion, thought, feeling, and associated physical responses. There is no such thing as a metaphysical feeling/learning.

Reference

Jackson, F. (1986). What Mary Didn't Know. *The Journal of Philosophy*, 83(5), 291–295.
<https://doi.org/10.2307/2026143>