

**teorema**

Vol. XXXIX/2, 2020, pp. 87-103

ISSN: 0210-1602

[BIBLID 0210-1602 (2020) 39:2; pp. 87-103]

## A Priori Physicalism and the Knowledge Argument

Tufan Kıymaz

### RESUMEN

Defiendo el argumento del conocimiento de Frank Jackson de la objeción de corte fisicista *a priori* de que, por todo lo que sabemos, la tesis de la deducibilidad (es decir, la tesis de que el conocimiento fenoménico es deducible en principio del conocimiento físico/funcional) es verdadera. Analizo la fuerza de la intuición ordinaria contra la tesis de la deducibilidad y la plausibilidad de los intentos que hay en la literatura por defenderla. Defiendo que los fisicistas *a priori* fracasan tanto a la hora librarse de la carga de la prueba que les obliga a convencernos de que la tesis de la deducibilidad es plausible, como a la hora de proporcionar una situación inteligible en la que el conocimiento fenoménico puede ser deducido del conocimiento físico/funcional.

PALABRAS CLAVE: *fisicismo a priori, deducibilidad, argumento del conocimiento, conciencia fenoménica.*

### ABSTRACT

I defend Frank Jackson's knowledge argument against the *a priori* physicalist objection that for all we know, the deducibility thesis, the thesis that phenomenal knowledge is in principle deducible from physical/functional knowledge, is true. I discuss the force of the common intuition against the deducibility thesis and the plausibility of the attempts in the literature in defense of the deducibility thesis. I argue that *a priori* physicalists both fail to remove the burden of proof from their shoulders to convince us that deducibility thesis is plausible and fail to offer an intelligible scenario in which phenomenal knowledge is possibly deduced from physical/phenomenal knowledge.

KEYWORDS: *A Priori Physicalism, Deducibility, Knowledge Argument, Phenomenal Consciousness.*

Frank Jackson's (1982, 1986) Mary is a brilliant neuro-scientist who has never left her black-and-white room, in which she has never seen colors. She acquires complete physical knowledge, knowledge expressible in physical/functional terms<sup>1</sup>, about human color vision from books and a black-and-white television. She is also cognitively ideal, that is, she is capable of deducing any further knowledge that is entailed by what she already knows, and she never ignores a relevant piece of information.

Jackson writes:

She knows all the physical facts about us and our environment, in a wide sense of ‘physical’ which includes everything in *completed* physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles. If physicalism is true, she knows all there is to know. For to suppose otherwise is to suppose that there is more to know than every physical fact, and that is what physicalism denies [Jackson (1986), p. 291; emphasis in the original].

But one day she leaves the room, sees a red tomato, and learns what it is like to see red. Since she learns something new, her pre-release knowledge, complete physical knowledge, was not complete knowledge, and hence, there are nonphysical truths and physicalism is false.

Jackson puts his argument more formally as follows:

- K1. Mary (before her release) knows everything physical there is to know about [the color experiences of] other people.
- K2. Mary (before her release) does not know everything there is to know about [the color experiences of] other people (because she *learns* something about them on her release).
- K3. Therefore, there are truths about [the color experiences of] other people (and herself) which escape the physicalist story. [Jackson (1986), p. 293; emphasis in the original].

As an objection, some philosophers argue that K2 is false, or, we don’t have sufficient epistemic reason to believe that K2 is true. Some of the objectors admit that Mary has an a-ha moment and she experiences a sense of discovery when she sees red for the first time, but they argue that she does not learn a new fact. Her sense of discovery, according to these philosophers, can be explained by her gaining new cognitive abilities,<sup>2</sup> her becoming acquainted with a property,<sup>3</sup> or her acquiring a new phenomenal concept,<sup>4</sup> none of which amounts to learning a new truth.

A priori physicalists, like Churchland [(1985), p. 25-28] and Dennett [(2004), pp. 60-61; 2007], who deny the existence of an explanatory gap,<sup>5</sup> on the other hand, deny that Mary would have an a-ha moment. They argue that the intuition that Mary will have an a-ha moment is *prima facie* powerful only because of our lack of understanding of what complete physical knowledge would be. They maintain that it is question begging to just assume that a person with complete physical knowledge would

experience a sense of discovery when she sees red for the first time. In this paper, I defend the knowledge argument against this a priori physicalist objection.

## I. THE KNOWLEDGE INTUITION

Stoljar and Nagasawa [(2004), pp. 2-3] identify the intuition behind the knowledge argument as follows:

*The Knowledge Intuition:* No amount of physical knowledge is by itself sufficient for phenomenal knowledge.

Alter (2007), based on Stoljar and Nagasawa's observations, identifies three claims, at least one of which must be rejected in order to reject the knowledge intuition:

1. *The Complete-Knowledge Claim:* Before leaving the room, Mary knows everything physical.
2. *The Learning Claim:* Upon leaving, she learns something.
3. *The Non-Deducibility Claim:* If 1 and 2 are true, then what Mary learns when she leaves the room cannot be a priori deduced (deduced by reason alone, without empirical investigation) from the complete physical truth [Alter (2007), p. 397].

Note that from 1 and 2, one might infer not only that Mary fails to a priori deduce what it is like to see red from physical knowledge, but also that Mary fails to *imagine* what it is like to see red based on the physical description of the red-seeing phenomenal state. Even though pre-release Mary couldn't a priori deduce phenomenal knowledge from physical knowledge, she would know what it is like to see red if she accurately imagined what it is like to see red. Consider the following intuition:

*Imagination Intuition:* No phenomenal experience can be accurately imagined based solely on physical knowledge about that experience.

Some philosophers, such as Churchland [(1985), p. 25], Maloney [(1985), p. 36] and Dennett (2007) interpret the knowledge argument such that it

rests on the imagination intuition rather than the knowledge intuition (as analyzed by Alter (2007)). However, Jackson explicitly states that imagination is irrelevant to his argument. He writes:

The knowledge argument does not rest on the dubious claim that logically you cannot imagine what sensing red is like unless you have sensed red. Powers of imagination are not to the point. The contention about Mary is not that despite her fantastic grasp of neurophysiology and everything else physical, she could not imagine what it is like to sense red; it is that, as a matter of fact, she would not know [Jackson (1986), p. 292].

Jackson's argument is about a priori deducibility of phenomenal knowledge and not about imaginability of phenomenal experiences.

## II. WHAT MARY KNOWS

Jackson [(1986), pp. 292-293] states that Mary "knows everything physical" and "all the physical facts about us and our environment." One might object that simply assuming that Mary's knowledge includes all physical facts but not phenomenal facts is question-begging against the physicalists, since if physicalism is true, phenomenal facts are among the physical facts. We can avoid this problem by reformulating the knowledge argument as Nida-Rümelin (2015) does. In Nida-Rümelin's version of the knowledge argument, the first premise is not that Mary knows every physical fact, but that she has complete physical knowledge, which is knowledge expressible in physical/functional terms, and the claim that Mary knows every physical fact is supported by this premise. Jackson does not make this distinction and interpret physicalism both as the theory that all facts are physical facts and as the theory that "all (correct) information is physical information" [Jackson (1982), p. 127].

But, still, we need more clarification. Physical knowledge is knowledge that is expressible in physical (and functional) terms, but what are these physical terms? Are they only the fundamental physical terms such as "spin," "charge," and "mass," or do they include terms that refer to macro-level physical entities and phenomena, such as "molecule," "neural network," "organism," and "tomato?"

As we have seen, Jackson states that Mary's physical knowledge "includes everything in *completed* physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles" [Jackson (1986), p. 291;

emphasis in the original]. According to this description, Mary's knowledge is not merely micro-physical.

Horgan (1984) objects that simply assuming that Mary's complete physical knowledge is knowledge of all physical facts (but not all phenomenal facts) is question begging, and Chalmers (2004), as a reply, argues that "physical knowledge" should be understood *narrowly*, such that it is knowledge expressible in the language of micro-physical theory, in which case what refutes physicalism is not that phenomenal facts are nonphysical (since, according to this sense of "physical," they are nonphysical if they are not micro-physical), but that phenomenal facts are not metaphysically necessitated by (micro-) physical facts. However, there is a problem with Chalmers's reply. If we understand physical knowledge in this narrow sense, then it is possible that, even if physicalism is true, Mary has complete physical knowledge, but she still learns something new when she sees red, simply because she hasn't done the relevant deductions yet. So, we should add the assumption that, along with her complete (narrow) physical knowledge, she also knows every truth (concerning phenomenology of color experience) that is deducible from her (narrow) physical knowledge. But, if the assumption that complete macro-level physical knowledge does not include phenomenal knowledge is question begging, so is the assumption that complete micro-level physical knowledge *and* the physical knowledge of every truth deducible from complete micro-level physical knowledge do not include phenomenal knowledge. So, if the original formulation by Jackson is question begging, so is Chalmers's proposal.

However, I don't think the claim that Mary has complete broad (both micro-level and macro-level) physical knowledge begs the question against the physicalist. The claim that macro-physical knowledge does not include phenomenal knowledge is not simply assumed, but it is supported by the learning claim. As I understand the knowledge argument, the argument is that if physicalism is true, then complete broad physical knowledge must *include* phenomenal knowledge,<sup>6</sup> but we have the strong intuition that even though Mary has complete broad physical knowledge, she still learns something new when she sees red, which implies that her complete broad physical knowledge is not complete knowledge about color vision. The important point here is that broad physical knowledge is, like narrow physical knowledge, objective knowledge that can be learned discursively. Based on Mary's case, we tend to believe that complete broad physical knowledge cannot contain phenomenal knowledge,

because, intuitively, phenomenal knowledge of what it is like to see red cannot be learned discursively from books and lectures. So, in Jackson's argument, the claim that phenomenal knowledge is not part of (or cannot be deduced from) complete broad physical knowledge is supported by the learning claim, and there is nothing question begging in the premise that Mary has complete broad physical knowledge.

Another important feature of Mary's physical knowledge is that it is *general* knowledge. She is not omniscient about every *particular* fact that is expressible in physical terms. The *general* physical facts that she knows include the correct scientific account of color and color vision, facts such as that there are red objects (objects that reflect light with wavelength of about 650 nm if exposed to visible light under normal conditions), that there are people who see these objects under normal conditions, and so on. According to Jackson, what Mary learns when she sees red for the first time is also general knowledge about human color experience and if physicalism were true every general truth about human color experience would be deducible from general physical knowledge.

To sum up, according to Jackson, Mary's pre-release physical knowledge is general objective knowledge that can be expressed in broadly physical (both micro-physical and macro-physical) terms.

### III. WHAT MARY SHOULD HAVE KNOWN

Consider Jackson's following remarks about Fred, who has a color experience that we do not have.

If Physicalism were true, enough physical information about Fred would obviate any need to extrapolate or to perform special feats of imagination or understanding in order to know all about his special color experience. *The information would already be in our possession.* But it clearly isn't. That was the nub of the argument. [Jackson (1982), p. 132, emphasis in the original]

Jackson maintains that if phenomenal facts were physical facts, then no "special feats of imagination" would be required to derive phenomenal knowledge from physical knowledge. Furthermore, Jackson claims that "special feats of understanding" would not be required either. According to the above quote, if physicalism were true, then a person who has complete physical knowledge would *a fortiori* be in possession of the knowledge of any phenomenal truth. So, Jackson claims that if physicalism were true, phenomenal knowledge would be contained in complete

physical knowledge: It cannot be the case that Mary knows all physical truths but does not know a phenomenal truth. So, if Mary's physical knowledge were expressed as a long conjunction, then phenomenal knowledge of what it is like to see red would be a conjunct, and phenomenal knowledge would be deducible from physical knowledge (in conjunctive normal form) simply by conjunction elimination. And, then, Jackson argues that the fact that Mary learns something new when she sees a red tomato for the first time disproves physicalism, since her learning entails that a truth about what it is like to see red is not a priori deducible from (and hence not contained in) her pre-release knowledge.

Jackson does not explicitly argue for his claim that post-release Mary would have an a-ha moment and learn something new about the phenomenology of red-seeing experience; he takes this to be intuitively true. He writes that "it seems just obvious that she will learn something about the world and our visual experience of it" when she sees colors for the first time [(1982), p. 130]. But a priori physicalists disagree. Dennett, for example, calls this the "She'll be surprised, dammit!" intuition and argues that it is unwarranted [Dennett (2007), p. 17]. One can claim that Jackson's intuition is false based on either of these two theses:

*The Knowability Thesis:* It is possible to have phenomenal knowledge based on physical knowledge.

*The Deducibility Thesis:* It is possible to a priori deduce phenomenal knowledge from physical knowledge.

These two theses are not equivalent. If the deducibility thesis is true, so is the knowability thesis, but not vice versa. One might claim that the knowability thesis is true since Mary can have phenomenal knowledge based on physical knowledge because she can imagine a phenomenal state solely based on the physical description of that state.<sup>7</sup> But, clearly, this can be the case while the deducibility thesis is false. As we will see below, philosophers who reject Jackson's intuition about Mary's post-release knowledge tend to confuse the knowability thesis with the deducibility thesis and fail to address Jackson's actual argument. When Jackson claims that Mary learns something upon seeing red for the first time, he rejects the deducibility thesis rather than the broader knowability thesis, since, as we have seen, according to Jackson, if physicalism is true, then phenomenal knowledge is contained in physical knowledge and can be

deduced from it by conjunction elimination without “any need to extrapolate or to perform special feats of imagination or understanding.”

But, before focusing on the objections to Jackson, let us ask why one would think that the deducibility thesis is false (and the learning claim is true), aside from claiming that it is intuitively true that Mary would be surprised upon seeing red for the first time.

One line of consideration in favor of the falsity of the deducibility thesis derives from Nagel’s seminal paper “What Is It Like to Be a Bat?” (1974). Let’s suppose Mary has complete physical knowledge about bats, and she wonders what it is like to experience echolocation (to see with her ears, so to speak). Because this experience is too alien to her, she is not capable of undergoing (or accurately imagining) the phenomenal experience of echolocation. And, arguably, if one is not capable of experiencing (or imagining) a certain phenomenal state, then one is not capable of acquiring the phenomenal knowledge of what it is like to be in that state. So, Mary *can* have complete physical knowledge about echolocation but she *cannot* have the phenomenal knowledge of echolocation, which means that phenomenal knowledge of echolocation cannot be a priori deduced from (or contained in) complete physical knowledge.

This is a common-sense argument against the deducibility thesis, but one can argue that it begs the question against the deducibility thesis to insist that it is definitional that if one is not capable of experiencing (or accurately imagining being in) a certain phenomenal state, then one is not capable of acquiring the phenomenal knowledge of that state.

Here is a second argument against the deducibility thesis. Let  $c_R$  be the fine-grained proposition expressed by post-release Mary in Jackson’s argument as “*this* is what it is like to see red” (whereby she expresses her newly acquired phenomenal knowledge). Suppose, for reductio, that the deducibility thesis is true and pre-release Mary deduces what it is like to see red from her physical knowledge. Let  $p$  be the conjunction of all true physical/functional statements that pre-release Mary knows. Then some conjunct of  $p$  expresses  $c_R$ . Let’s call this conjunct “ $p_R$ .” So,  $p_R = c_R$  and pre-release Mary knows  $p_R$ .  $p_R$ , however, has the form *what it is like to see red is  $\varphi$* , but it does not include a demonstrative (or phenomenal) concept as  $c_R$  apparently does;  $p_R$  most likely includes information about the human visual system, the functional/causal facts about red-seeing experience, the physics of red light and its interaction with the retina, and so on. Therefore,  $\varphi$  is a long list of physical/functional properties of the red-seeing experience.

How long a description is  $\varphi$ ? Dennett writes:



Are we really so sure that what it is like to see red or blue can't be conveyed to one who has never seen colors in *a few million or billion words*? ... Remember, Mary knows everything about color that can be learned by physical science, and she presumably has the attention span and powers of comprehension required to handle *10 billion words* on what it is like to see red [Dennett (2007), p. 21, my emphasis]

So,  $\varphi$  might be a description of what it is like to see red that involves maybe billions of words. In fact, there is no word limit as long as it is finite, but the idea is that  $\varphi$  is so long and detailed that our intuition about the inexpressibility of phenomenal experience is not reliable anymore, given the possibility of such a long and complex description that is incomprehensible for us with our cognitive limitations.

Now, suppose at  $t_1$  pre-release Mary deduces that what it is like to see red is  $\varphi$  from her complete physical knowledge, but then the next day, at  $t_2$ , she hits her head in the room and as a result of a mild concussion she misremembers one fact about the causal/functional role of the red-seeing experience. Say, instead of the fact that the red-seeing experience is normally triggered by observation of light with 650 nm wavelength, she, after  $t_2$ , believes that the red-seeing experience is normally triggered by observation of light with 560 nm wavelength. Now, she believes that what it is like to see red is  $\varphi'$ , where  $\varphi'$  mentions all the properties mentioned in  $\varphi$  except one.

So, at  $t_1$  pre-release Mary knows what it is like to see red since she deduces it from her complete physical knowledge, but at  $t_2$ , even though she still believes that she knows what it is like to see red, she is wrong because she misremembers a fact about the wavelength of red light. Now, the intuition pump: It is very hard to see how misremembering just one detail about the physical/functional facts about the red-seeing experience among maybe millions of truths obliterates the phenomenal knowledge of what it is like to see red. In other words, it follows from Dennett's account that phenomenal knowledge is a kind of knowledge that can be highly sensitive to very small mistakes about very complex descriptions, which is intuitively not the case. Furthermore, Mary's belief that what it is like to see red is  $\varphi'$  is still a phenomenal belief,<sup>8</sup> that is, it is a belief that would normally be expressible as "*this* is what it is like to see red" where "this" denotes a phenomenal state. But, what phenomenal state does  $\varphi'$  describe? There is no such phenomenal state in human experience that is accurately described by  $\varphi'$ . But, how can we make sense of Mary's false belief at  $t_2$ ? Yellow light has 560 nm wavelength, so, does

she mistakenly believe that what it is like to see red is  $x$  where  $x$  denotes the phenomenal state that we normally experience when we see yellow? But, every other physical, neurological, functional detail in  $\varphi'$  is about the red-seeing experience, and many of them are not true for yellow-seeing experience.

Phenomenal knowledge does not seem to us to be expressible in complex physical terms because it is relatively simple. When I look at a particular tree and think “this is a tree,” on the other hand, I don’t think that the reference of “this” in principle escapes any description in physical/functional terms. It is quite possible that everything intrinsic about the tree I am attending is expressible in physical/functional terms, at least in principle. I can describe *this*, the tree in front of me, by giving you information about its shape, the material it is made out of, its spatio-temporal location, and so on. The description of this tree would provide a list of its properties. But, when I look at a red object, attend to my color experience, and think “*this* is what it is like to see red,” it seems to me that the reference of “this” cannot be captured by any objective description, because I cannot see how a list of physical/functional properties can be its complete description while the thing that I am trying to describe, *from my first person perspective*, has only one property, in fact, the reference of “this” is that very property that I want to describe: the phenomenal character of the red-seeing experience. The phenomenal character might also seem to play some functional/causal role, but a list of those roles would not be a description of *this* distinct phenomenal character, since *this* is what plays those functional/causal roles. I am not claiming that this line of thinking proves that phenomenal character is not reducible to physical/functional properties, but I do think, contrary to Dennett’s claim, that the burden of proof is on the proponents of the deducibility thesis to give us a story about how a physical/functional description, even if it is billions of words long, can capture the phenomenal character of an experience.

#### IV. WHAT DEDUCTION IS NOT

There are indeed some suggestions in the literature about how Jackson’s intuition that Mary learns something when she sees red for the first time might be false. I will consider three of such accounts, each of which rests on some misconception about deducibility or the knowledge argument.

1. Let's start with a misunderstanding about Jackson's argument. Maloney [(1985), p. 34] argues that Mary can know what it is like to see red while she is in the room since "if ... Mary does understand *all* that there is to know regarding the physical nature of colour vision, she would be in a position to imagine what colour vision would be like." Churchland [(1985)], like Maloney, interprets the knowledge argument to be about Mary's inability to imagine phenomenal redness based on her physical knowledge, and argues that it is epistemically possible that Mary could imagine phenomenal redness based on complete physical knowledge. But, doing so, he summarily dismisses Jackson's argument, even though he agrees that the deducibility thesis is false.<sup>9</sup> Similarly, Dennett writes "we are told that Mary in her cell can't imagine what it's like to experience red, try as she might" [Dennett (2007), p. 23].

However, as we have seen above, Jackson explicitly states that imagining is irrelevant to his argument and he writes "the knowledge argument claims that Mary would not know what the relevant experience is like. What she could imagine is another matter" [Jackson (1986), p. 295]. The objection from imaginability fails to address Jackson's argument.

2. Dennett argues that it is possible that Mary, before she leaves the room, would be able to correctly identify a random novel color experience by "noting some salient and specific reaction that her brain would have only for yellow or only for red" [Dennett (2004), p. 61], and therefore, if she were presented with a blue banana after her release, she would be able to say "Hey! You tried to trick me! Bananas are yellow, but this one is blue!" [Dennett (2004), p. 60].

Dennett argues that post-release Mary would not be surprised; however, the scenario that he considers has no bearing on the deducibility thesis, since the reason why Mary is not surprised in this scenario is not that she has deduced the phenomenal knowledge from physical knowledge. It is possible that the deducibility thesis is false while this scenario is true, because Dennett's Mary identifies the phenomenal property simply by matching her current phenomenal experience with her current neural state, without making a *deduction* from her general physical knowledge.

Furthermore, as Robinson [(1993), p. 75] points out, Dennett confuses general and particular knowledge, since he assumes that, at any moment, Mary has complete physical knowledge about the particular neural state that she is in at that moment. As I've mentioned in section II, Mary does not have complete knowledge about every particular phys-

ical fact. What people experience when they see red is a general fact about human color vision and Mary is supposed to deduce the knowledge of that fact from her general physical knowledge. On a similar note, Jacqueline [(1995), p. 225] points out that, given Dennett's interpretation of Mary's physical knowledge, Mary would know that the banana is blue rather than yellow simply based on her particular physical knowledge about the wavelength of light that is reflected from the banana, which need not involve actually leaving the room and seeing it.

3. Dennett also specifically objects to Jackson's claim that phenomenal knowledge is *not deducible* from physical knowledge, and hence, *seemingly*, defends the deducibility thesis. I say "seemingly," because in his discussion of deducibility of phenomenal knowledge, he interprets *deducibility* in such a way that what he actually defends fails to be adequately similar to the deducibility thesis. In the scenario he proposes, Mary,<sup>10</sup> based on her complete physical knowledge, knows which neural state is the dispositional state of knowing what it is like to see red; let's call it *state B*. She puts herself in state B, and thereby comes to know what it is like to see red. State B does not involve actual or imagined experience of seeing red; however, it is the competence state that a color experience normally causes [Dennett (2007), pp. 28-9]. So, Mary would know what it is like to see red by putting herself not in the state of *experiencing* what it is like to see red, but in the state of *knowing* what it is like to see red.<sup>11</sup>

Dennett considers the following objection to his scenario: "What matters is whether Mary (or RoboMary) can *deduce* what it's like to see red from her complete physical knowledge, not whether one could use one's physical knowledge in some way or other to acquire knowledge of what it's like to see in color," and he replies: "So far as I can see, this objection presupposes an improbable and extravagant distinction between (pure?) deduction and other varieties of knowledgeable self-enlightenment" [Dennett (2007), p. 29]. He insists that, in his scenario, since Mary does not illicitly use her physical knowledge to experience or imagine what it is like to see red, her way of acquiring phenomenal knowledge can be regarded as deduction from physical knowledge.

But, Dennett's notion of deducibility is too liberal to be relevant to the knowledge argument. In his example, Mary knows what neural state she would be in if she knew what it is like to see red, she puts herself in that state, and she thereby comes to know what it is like to see red; however, this scenario can be true even if physicalism is false. Suppose we have immaterial souls and some kind of parallelism is true, and our men-

tal states are nomically but not metaphysically determined by physical states. Then, Mary would still acquire the phenomenal knowledge in question by putting herself into the neural state of knowing what it is like to see red. So, according to Dennett's notion of deducibility, the knowledge of *nonphysical* phenomenal properties that do not metaphysically supervene on but are just nomically determined by physical properties can be deduced from physical knowledge. So, Dennett's notion of deducibility would permit the deducibility of phenomenal knowledge from physical knowledge even if physicalism was false. Or, consider another scenario. Suppose the ability hypothesis<sup>12</sup> is true and to have the phenomenal knowledge of what it is like to see red is to have some cognitive abilities. Now, following Dennett's example, suppose Mary can put herself in any neural state. Then, Mary can put herself in a neural state in which she has the cognitive abilities that constitute the phenomenal knowledge of what it is like to see red. In this scenario, phenomenal knowledge is non-propositional, but it is still deducible from physical knowledge according to Dennett's understanding of deducibility. This, again, shows that what Dennett calls "deducibility" is very different from the kind of deducibility that is in question in our discussion of the knowledge argument.

One last very serious problem with Dennett's view is that his view allows the deducibility of false statements from physical truths. Suppose Mary knows which neural state corresponds to the belief that  $2+2=5$  and puts herself into that neural state. Now, it follows from Dennett's account of deducibility that Mary, in this scenario, deduces  $2+2=5$ , which is a necessarily false statement, from her true physical/functional beliefs about a neural state. I don't think this non-truth-preserving notion of deducibility has any place in our discussion on the relation between physical/functional knowledge and phenomenal knowledge.

## V. CONCLUSION

I have argued that the burden of proof is on the a priori physicalist who claims that pre-release Mary can deduce phenomenal knowledge of what it is like to see red from physical/functional information about color vision. One might, however, defend a more modest position and claim that we do not have good epistemic reasons to favor either the claim that phenomenal knowledge is deducible from physical/knowledge or the claim that it is not. Still, I think the considerations in section III, with the

failure of the a priori physicalist responses in section IV, do provide us good reason to think that phenomenal knowledge is not deducible, and hence we should also reject this more modest position.

*Department of Philosophy , H249  
Bilkent University  
Ankara 06800, Turkey  
E-mail: tufan.kiyamaz@bilkent.edu.tr*

#### NOTES

<sup>1</sup> Plus, of course, mathematical/logical terms, but I will not be explicitly mentioning this in the remainder of the paper.

<sup>2</sup> Lewis (2004) and Nemirow (1980, 2007) argue that Mary does not learn new factual/propositional knowledge when she experiences a new phenomenal state, but she gains some cognitive abilities, which constitute not knowledge-that but know-how. According to Lewis [(2004)], knowing what it is like to have a certain experience consists in the possession of the abilities to recognize, imagine, and remember the relevant experience.

<sup>3</sup> Earl Conee (1994), argues that Mary's new phenomenal knowledge is not factual knowledge, but nonfactual acquaintance knowledge. According to Conee (1994), p. 144, acquaintance knowledge consists in the most direct way of apprehending a thing, and this is not a type of knowledge that is peculiar solely to first-personal knowledge of phenomenal states. He states that "to come to know a property is to become acquainted with the property, just as to come to know a city is to become acquainted with the city, and to come to know a problem is to become acquainted with the problem" [Ibid., p. 140]. So, Mary's new knowledge does not pose a threat to physicalism, since it is nonfactual acquaintance knowledge and the phenomenal property that she becomes acquainted with can be a physical/functional property.

<sup>4</sup> According to the view Stoljar (2005) dubs "the phenomenal concept strategy," Mary's new phenomenal knowledge is about the physical facts that she already has physical knowledge about. Phenomenal knowledge is knowledge that involves phenomenal concepts, and (excluding miracles) phenomenal concepts can only be acquired by phenomenal experience. A phenomenal concept picks out its reference directly, and not via a contingent property of its reference. What a phenomenal concept picks out, according to this view, is a physical property. So, what Mary learns upon seeing a red tomato is not a new truth, but a new way of apprehending a truth that she already knows in physical terms. Some proponents of this view are Horgan (1984), Byrne (2002), Loar (2004), Levin (2007), Papineau (2002, 2007), Balog (2012).

<sup>5</sup> See Levine (1983), (2001).

<sup>6</sup> I'll defend this interpretation in the next section.

<sup>7</sup> See Kıymaz (2019) for an argument against physicalism that is based on pre-release Mary's supposed inability to imagine a phenomenal experience rather than her lack of phenomenal knowledge.

<sup>8</sup> Dennett does not discuss phenomenal beliefs but focuses only on phenomenal knowledge. But I think it would be acceptable for Dennett to talk about phenomenal beliefs since according to him all phenomenal knowledge is propositional knowledge.

<sup>9</sup> Churchland writes:

It is true, of course, that no sentence of the form "x is a sensation-of-red" will be deducible from premises restricted to the language of neuroscience. ... direct deducibility is an intolerably strong demand on reduction, and if this is all the objection [to physicalism] comes to, then there is no objection worth addressing. What the defender of emergent qualia must have in mind here, I think, is the claim that Mary could not even imagine what the relevant experience would be like, despite her exhaustive neuroscientific knowledge. [Churchland (1985), p. 25]

Churchland, like Jackson, rejects the deducibility thesis but he claims that physicalism is not committed to "direct deducibility" of phenomenal knowledge from physical knowledge.

<sup>10</sup> Dennett talks about *RoboMary* in his example, but in my discussion I will apply his example to Mary.

<sup>11</sup> I am not sure how someone can put herself in a given neural state, given that we don't exercise direct control on our neurons. Dennett offers a thought experiment involving *RoboMary* and writes that she "—being such a clever, indefatigable and nearly omniscient being— makes all the necessary adjustments and puts herself into state B" where state B is the state of having phenomenal knowledge. His point is that this scenario is possible (even if not practically possible), and this counts as deduction from physical knowledge, therefore, phenomenal knowledge is in principle deducible from physical knowledge.

<sup>12</sup> See note 2.

## REFERENCES

- ALTER, T. (2007), "The Knowledge Argument"; in *The Blackwell Companion to Consciousness*, edited by Max Velmans and Susan Schneider, Blackwell Publishing, 396-405.
- . (2008), "Phenomenal Knowledge Without experience"; in Edmond Wright (ed.), *The Case for Qualia*, Cambridge, Mass., MIT Press, pp. 247-267.
- BALOG, K. (2012), "In Defense of the Phenomenal Concept Strategy"; *Philosophy and Phenomenological Research* 84 (1), pp.1-23.
- BYRNE, A. (2002), "Something About Mary"; *Grazer Philosophische Studien* 62 (*Essays in Honour of the Philosophy of Terence Horgan*), pp. 123–140.

- CHALMERS, D. J. (2004), “Phenomenal Concepts and the Knowledge Argument”; in P. Ludlow, D. Stoljar and Y. Nagasawa (eds.), *There’s Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson’s Knowledge Argument*, Cambridge, Mass., MIT Press, pp. 269-298.
- CHURCHLAND, P. M. (1985), “Reduction, Qualia and The Direct Introspection of Brain States”; *Journal of Philosophy* 82 (January), pp. 8-28.
- CONEE, E. (1994), “Phenomenal Knowledge”; *Australasian Journal of Philosophy* 72 (2), pp. 136–50. doi:10.1080/00048409412345971.
- DENNETT, D. C. (2004), “Epiphenomenal Qualia?”; in P. Ludlow, D. Stoljar and Y. Nagasawa (eds.), *There’s Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson’s Knowledge Argument*, Cambridge, Mass., MIT Press.
- (2007), “What RoboMary Knows”; in Alter, T., and S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge*, Oxford, Oxford University Press.
- HORGAN, T. (1984), “Jackson on Physical Information and Qualia”; *Philosophical Quarterly*, 34, pp. 147-52.
- JACKSON, F. (1982). “Epiphenomenal Qualia”; *The Philosophical Quarterly* 32 (127), pp. 127–36. doi:10.2307/2960077.
- (1986), “What Mary Didn’t Know”; *The Journal of Philosophy* 83 (5), pp. 291–95, doi:10.2307/2026143.
- JACQUETTE, D. (1995), “The Blue Banana Trick: Dennett on Jackson’s Color Scientist”; *Theoria* 61 (3), pp. 217-30.
- KIYMAZ, T. (2019), “What Gary Couldn’t Imagine”; *Journal of Philosophical Research* 44, pp. 293-311.
- LEVIN, J., (2007), “What is a Phenomenal Concept?”; in Alter, T., and S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge*, Oxford, Oxford University Press, pp. 87–110.
- LEVINE, J. (1983), “Materialism and Qualia: The Explanatory Gap”; *Pacific Philosophical Quarterly* 64, pp. 354-61.
- (2001), *Purple Haze: The Puzzle of Consciousness*; Oxford, Oxford University Press.
- LEWIS, D. (2004), “What Experience Teaches”; in P. Ludlow, D. Stoljar and Y. Nagasawa (eds.), *There’s Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson’s Knowledge Argument*, Cambridge, Mass., MIT Press, pp. 77-104.
- LOAR, B. (2004), “Phenomenal States (Revised Version)”; in P. Ludlow, D. Stoljar and Y. Nagasawa (eds.), *There’s Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson’s Knowledge Argument*, Cambridge, Mass., MIT Press, pp. 219-240.
- MALONEY, J. C. (1985), “About Being a Bat”; *Australasian Journal of Philosophy* 63 (1), pp. 26-49.
- NAGEL, T. (1974), “What is it like to be a bat?” *Philosophical Review* 83 (October), pp. 435-50.
- NEMROW, L. (1980), “Review of Nagel’s Mortal Questions”; *Philosophical Review*, 89, pp. 473-477.



- (2007), “So, This is What It’s Like: A Defense of the Ability Hypothesis”; in Alter, T., and S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge*, Oxford, Oxford University Press, pp. 32-51.
- NIDA-RÜMELIN, M. (2015), “Qualia: The Knowledge Argument”; *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), Edward N. Zalta(ed.), forthcoming URL = <<http://plato.stanford.edu/archives/sum2015/entries/qualia-knowledge/>>.
- PAPINEAU, D. (2002), *Thinking about Consciousness*; Oxford, Clarendon Press.
- (2007). “Phenomenal and Perceptual Concepts”; in Alter, T., and S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge*, Oxford, Oxford University Press, pp. 111-144
- ROBINSON, H. M. (1993), “Dennett on the Knowledge Argument”; *Analysis* 53 (3), pp. 174–77.
- STOLJAR, D. (2005), “Physicalism and Phenomenal Concepts”; *Mind & Language* 20 (5), pp. 469–94; doi:10.1111/j.0268-1064.2005.00296.x.
- STOLJAR, D. and NAGASAWA, Y. (2003), “Introduction”; in P. Ludlow, D. Stoljar and Y. Nagasawa (eds.), *There’s Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson’s Knowledge Argument*, Cambridge, Mass., MIT Press, pp. 1-36.