

The Knowledge Argument: experience and acquaintance towards a world of colors

Abstract: the aim of the paper is to recall Frank Jackson's 1982 knowledge argument and some prominent objections that were brought against it. Is there a new path we can take in order to bring something new to the table? Is the debate on the argument and its powerful conclusion over?

The paper is divided in four sections. After a brief introduction which explains the structure of the knowledge argument, Section 1 considers the ability hypothesis by Lewis and Nemirow. Section 2 reflects on the possibility that Mary doesn't learn anything new, with references to Dennett and Churchland's objections. Section 3 considers another response, which is known as "phenomenal concept strategy". Finally, in Section 4 I advance a possible solution to the knowledge argument, recalling the acquaintance response given by philosophers such as Brie Gertler and, partially, David Chalmers.

Introduction

In 1982 Frank Jackson's famous paper, the philosopher presents a thought experiment that will influence the debate on the nature of consciousness for almost forty years. The idea presented in the script is the following: we are asked to imagine a brilliant neuroscientist, Mary, who achieves a complete knowledge about the human color vision. She masters chemistry, physics, biology, in sum all the empirical sciences which are concerned with sight and colors (e.g. she studies the different wavelengths of the spectrum and the reflection of the light on the retina). Furthermore, she is imprisoned in a very peculiar environment, that is to say in a completely black and white room. From there she carries on her studies on the topic via a black and white television and colorless books, which allow her to learn all the relevant physical facts on the human

color vision. After many years of agony and total despair from Mary, the scientists who kidnapped her decide that it was time to let her go. One question arises spontaneously: since Mary has never experienced the color red, as well as all the other variants of the spectrum, will she come to know anything new or not? The original answer given by Jackson in 1982 was affirmative: Mary, in fact, can finally experience what it is like (to use Nagel's terminology¹) to see the color red. To put it into more technical terms, she apprehends a new series of facts, the so-called "*qualia*", the subjective and ineffable experience everyone is familiar with in daily life (e.g. the smell of coffee, staring at a sunset, seeing the color green, having a pain in the left foot). Mary comes to know what it is like to see the color red (which is far different from seeing, for example, the color blue) and what the other people experience when they see that very color.

Where does this original line of thought lead us to? The conclusion of the argument is that since our famous neuroscientist knew, inside her personal black and white room, all there was to know from a physical point of view on the human color vision, what she learns *after* her release has got to be a new piece of information that is not physical (otherwise, she would have been able to deduce it when locked in her room).

This suggests that our reality can't be fully grasped using only the empirical sciences and, most importantly, a completely materialist (or physicalist²) explanation of everything: if we are ready to accept that Mary learns something she couldn't come to know in her room, then we have to recognize that qualia are non-physical and knowable only thanks to experience. Physicalism, the idea that reality can be explained only in physical terms and that everything existing has got to be physical (consciousness and qualia too), is then false.

The direct consequence of the knowledge argument, and this is the main reason behind the many attempts to undermine it from the moment it was published in 1982, is that we should embrace a form of property dualism (the "epiphenomenalism" Jackson talks about in the second part of his paper) and accept that our

1. See Nagel (1974).

2. During the period of the Vienna's Circle (1930's) there was a difference between "materialism" and "physicalism". The first term defined a metaphysical doctrine, whereas the second one referred to a linguistic one.

world can't be fully reduced using a physicalist formula. Despite of the intuitive appeal though, Jackson himself rejected his own argument in 1996³ due to a heavy line of objections, some of which were tremendous for the structure of the argument (a few will be discussed in the next sections below). Now he's joined the physicalist crew and sustains a position which is defined as "representationalism", which roughly states that qualia are objective properties of our represented experience (such as the color, the form and all the other physical proprieties of the case). The reasons behind Jackson's change of mind and more details on his current position can be found in Jackson (2003) and (2007), but he recently presented another paper containing more information on the general idea behind representationalism (2019).

Now that we have seen how the knowledge argument works, it is time to face some powerful objections which have tried to prove it wrong (not successfully, since the argument is still widely discussed nowadays). The following sections will focus on the most prominent objections to the argument, with some replies that are found in the literature. In the final section of the paper, I sketch a possible interpretation of the knowledge argument, which maintains that it is both valid and correct, appealing to the notion of acquaintance and accepting the intended conclusion of the argument (with particular reference to Gertler 2019).

Before doing so, I would like to list the premises of the argument, for a better understanding of the attacks from the physicalists party. Another compelling reason for doing so is clarity since many philosophers focused on different premises of the argument and this may help throughout the various sections. Here's the structure of the argument, even though many stipulations are valid and acceptable:

The Knowledge Argument:

(1) Mary knows all the physical facts concerning the human color vision before her release.

3. The shocking change of mind by Jackson was even reported on the Philosophical News Service.

(2) Mary does not know everything there is to know about human color vision, as she learns something new (what is like to see red) *after* her release.

(3) There are facts about the world which are non-physical (i.e. qualia).

Conclusion: physicalism is false.

Section 1. First objection: Mary learns an ability

The knowledge argument encountered a lot of resistance in the past years, especially the ten which followed from its publication. One of the first objections that were raised against the argument was brought by David Lewis (1983, 1988 and 1990) and Laurence Nemirow (1980, 1990 and 2007). The main idea behind this objection is that the second premise (2) is false. In fact, what Mary comes to know after being released from the black and white room isn't a quale (the singular for "qualia") or something non-physical, but an ability. What kind of ability are the authors referring to? Here are two passages, respectively from Lewis (1983) and (1990):

...knowing what it is like is the possession of abilities: abilities to recognize, abilities to imagine, abilities to predict one's behavior by imaginative experiments (Lewis 1983, 131).

The Ability Hypothesis says that knowing what an experience is like just *is* the possession of these abilities to remember, imagine, and recognize. ... It isn't knowing-that. It's knowing-how (Lewis 1990, 516).

It's important to stress out that Lewis and Nemirow intend the information acquired by Mary as a particular type of knowledge, that is to say "knowledge-how". A few examples: riding a bike, learning how to swim, comprehend how to solve a math problem. If the knowledge argument is sound, Mary should learn a new fact, which is considered in literature as "knowledge-that". An example can be the fact that you may know *that* the atomic number of the element gold is 79, or *that* water is H₂O, or once again *that* Trump is the actual president of USA. Once we concede that Mary learns what is like to see red (a phenomenological fact), her knowledge must be identified as a type of knowledge-that, which then leads her to an epistemic progress. In order to save physicalism, Lewis and Nemirow deny this step and concede instead that Mary comes to learn only a new piece of knowledge-how, with all the abilities listed before by Lewis.

The key point is that the thought experiment proposed by Jackson doesn't have to deal with non-physical stuff or properties that the actual physics can't categorize, but it should instead identify the new knowledge obtained by Mary with an ability concerning the color red. Once she is able to experience a new world full of colors outside of her boring environment, she can finally learn to imagine, recognize and discriminate the experience of red from the other ones (like the experience of the quale of yellow). Given that Lewis and Nemirow both share physicalists intuitions, their attempt to clarify the mysteries behind the knowledge argument have to be seen as a way to undermine it at its base, denying of course the original dualist conclusion.

Reply: an interesting answer can be found in Alter and Howell's "Dialogue con Consciousness", where are presented a couple of counterexamples to the ability hypothesis by Lewis and Nemirow. In the text, one of the two protagonists, Ponens, imagines the scene described by Jackson, but with a little variation. When Mary is confronted with a rose (the original paper mentions a ripe tomato), she surely comes to know what is like to see red. But once we take away the object from her, it's not clear whether she can (like it is proposed

by Alter and Howell) imagine that experience with the exact same hue she just saw. She could have seen a vermillion-like rose or a cherry-like one, without then being able to recall the precise experience she had. This brings us to the conclusion that what is like can't be *precisely* identical to an ability, whether it requires imaging, thinking, or scrutinizing.

A second case is described one page later in the same volume. This time Mary gets brain-damaged and loses the ability to remember (a sort of amnesia, in a sense) certain experiences. Here's the cited passage:

Anyway, the accident causes brain damage, and as a result her imaginative capacities become severely limited. [...] She knows what it's like to see red while staring at the rose, but the moment she turns away she can't imagine seeing red. So, while staring at the rose, she has phenomenal knowledge—she knows what is like to see red—without the corresponding imaginative ability. And the same reasoning applies to sorting abilities. (pp. 50-51)

In 1996, a few years after Lewis and Nemirow comments on the knowledge argument, a milestone of the philosophy of mind was published by David Chalmers. In "The Conscious Mind", Chalmers gives further support to the dualist conclusion of Jackson's argument. In a very precise and small passage, Chalmers argues that Mary surely learns an ability: she can imagine red, recognize it when she has the relevant experience and so on. But that's not all that she learns. Our intuition is that Mary makes a discover on the world that surrounds her. She comes to know something she didn't' know before and that she couldn't have known only by studying on her books. Chalmers's words explicit this concept perfectly:

No doubt Mary does gain some abilities when she first experiences red, as she gains some abilities when she learns to ride a bicycle. But it certainly seems that she learns something else: some facts about the nature of experience. For all she knew before, the experience of red things might have been like this, or

it might have been like that, or it might even have been like nothing at all. But now she knows that it is like *this*. She has narrowed down the space of epistemic possibilities. No such new knowledge comes along when an omniscient mechanic learns to ride a bicycle. (pp. 144-145).

It seems to me pretty evident that the answer provided by Lewis, Nemirow and all other philosophers who embrace the ability hypothesis is insufficient to explain what is like and its nature. Mary surely gains new knowledge-how and possesses a new set of abilities when she first sees a red object, but in order to get a grip on her epistemic progress we should look for another type of resolution. Section 2 considers another popular response, which focuses on the first premise of the argument.

Section 2. Second objection: Mary doesn't learn anything new

Another standard way to reject the conclusion of the knowledge argument is to deny that Mary actually learns something new after being released. This strategy focuses once again on the second premise of the argument and it is adopted by philosophers such as Dennett (1991 and 2007) and Churchland (1989). In "Epiphenomenal Qualia?", Dennett claims that Jackson's thought experiment doesn't work, as he can easily imagine a scenario in which Mary can figure out what is like to see red in her room. Once she is set free, she doesn't get the usual "aha!", the signal that she learnt something new (the quale of red). Instead, Mary knows exactly what it would have been like to undergo the qualia of a red object, thus showing that there are no non-physical qualia or facts that aren't physically explainable. The passage I am referring to states that Mary, instead of the usual rose/ripe tomato she gets in contact with, sees for the first time a blue banana⁴. The kidnappers tell her that the banana is yellow, but Mary quickly answers that it can't be the case, as she says that the banana is blue, where this information was available to her way before her release.

This would show that our brilliant neuroscientist doesn't gain any new knowledge on qualia, as she is capable of deducing that the banana is blue without appealing to experience or non-physical qualia. Mary doesn't get surprised of her discovery, as nothing new comes to light. Furthermore, Dennett replies to those who may be skeptical or even discontent about his proposal, saying that:

Surely, I have cheated, you think. I must be hiding some impossibility behind the veil of Mary's remarks. Can you prove it? My point is not that my way of telling the rest of the story proves that Mary doesn't learn anything, but that that the usual way of imagining the story doesn't prove that she does. It doesn't prove anything; it simply plumps the intuition that she does ("it seems just obvious") by lulling you into imagining something other than what the premises require. (pp 60-61)

Reply: How could a defender of the argument respond to these claims? A possibility is found in Torin Alter (2017), when the author says "She could use a brain scanner to examine her own brain processes while she views the banana; and she could notice that they correspond to brain processes that typically occur when people see blue objects". Alter insists that Dennett doesn't appeal to the correct source of information, because his scenario takes for granted that physical information is the key for Mary to succeed. Instead, as Alter sustains, Mary has got to refer to experience of seeing blue, otherwise she wouldn't fully understand what is like (see Alter 2008).

To me, the Mary described by Dennett couldn't have fully understood the difference between phenomenal yellow and blue: inside her room she can't have access to colors and having experience of both of them is necessary to distinguish between the two different hues. In other words, Mary has got to experience what is like to see blue in order to make the claim stressed by Dennett, something which unfortunately can't do only via the physical information. What's more, if she had been really capable of understanding what is like with

her books and her monitor, she wouldn't have needed to go outside and experience the blue banana in its phenomenal regards. What Dennett proposes as a counterexample to Jackson's argument is an attempt to undermine the thought experiment without analyzing the important role that experience has, and, furthermore, without a clear explanation of how Mary would have described what is like inside her room: no analysis, empirical investigation or extra research is brought to claim that Mary, in guise of superb scientist, can deduce what is like to see red in her room. Nothing could allow her to understand what is like, apart from having the direct experience of that specific quale. I'll return on this issue in Section 4.

Let's consider another response, which is close to Dennett's approach. Patricia Churchland and her husband Paul both share a passion for eliminativism and in their writings (fortunately) don't mention any blue banana trick. Strictly speaking, they both hold that our intuition that consciousness is something more that science can't explain in physical terms will vanish, as neurosciences make progress and get more information on the brain. Similar to the case of phlogiston in the XVII century, our belief in separate mental states and non-physical qualia will fade away once science will fully comprehend our brain and the phenomenon of conscious experience. But what about Mary? Paul Churchland (1989) tries to undermine the argument from its very first premise, which concerns the quantity of information Mary has access to before knowing what is like. Churchland insists that

“She thus lacks an appropriate portioned activation vector space across those neurons, and therefore has no representation, at that site, of the full range of sensory coding vectors that might someday come from the retina and the LGN. In other words, there is something physical about persons (their color sensations= their coding vectors in their visual pathways) ...”

The objection is pretty straightforward: if Mary hasn't access to the complex brain process described by Churchland, the very first premise and the entire knowledge argument are doomed. Mary doesn't know all there's to know about our color vision: she misses a crucial part that could help her to finally get a grip on what is like to see red. If Churchland is right, the argument is invalid, and physicalism is safe from the knowledge argument attack.

Here's another passage from Churchland, a few pages onward:

If we write a deliberately nonequivocal form of Jackson's argument, one that quantifies appropriately over all of the relevant forms of knowledge, then the first premise must also certainly be false under the conditions of his own story. (*ibid.*)

Churchland's hope is that neuroscience will break the wall between the first and the third person, giving us a complete physical model that doesn't have to appeal to dualism or non-physical qualia. Maybe it's time to let go our dualist intuitions to leave space to concrete sciences, which could eventually enlighten our poor knowledge on consciousness. Hempel's dilemma⁴ isn't a challenge to Churchland, but maybe there are some replies worth considering, before giving up the knowledge argument and convert ourselves to physicalism.

I didn't hear no bell.

Reply: I can't deny that many times the conclusion of the knowledge arguments strikes me as absurd. Why on earth should we rely on a thought experiment which doesn't give us any empirical proof of the existence of non-physical qualia? The reason is the following: the limit of Churchland's objection to the knowledge argument (which is of course far more complex and long than the version I have presented in this paper) is that, as a materialist, he doesn't treat qualia with justice. I agree with him that neurosciences are making

4. See Hempel (1969). Roughly, the idea is that the actual physics isn't complete, whereas the future one can't even be imagined. The definition of physicalism seems to be in trouble, as physics keeps on changing throughout the decades.

progress and that one day we may even come close to localize for sure consciousness in our brain. But that's not all there's to say about qualia and their story.

The Newton many philosophers are waiting⁵ is nowhere to be seen and subjective experience seems to escape the physicalist story, thanks to works by Chalmers and many others⁶. I find Churchland's behaviour way too overconfident on qualia and incapable of giving a concrete explanation of the subjective realm. Sure, dualists are not in a better spot than eliminativists, as they have to deal with mental causation and the mental, which seems ontologically unnecessary to the economy of the universe and the philosophy of mind. But honestly the intuition behind the knowledge argument and the other arguments that support anti-materialism are strong and well-motivated, thus we can't refute dualism because neurosciences pretend it to be false and illogical (while it is compatible, in some versions, even with the actual physics untouched⁷).

Concerning the attack on the first premise (I haven't forgotten about it), I am with Goff (2019) when he says that we intend Mary's knowledge as absolute, while we should accept her limits and consider her powerful knowledge as regarding colors and human vision. She doesn't have to know a theory of everything and categorize every aspect of reality without committing a single mistake: Jackson's idea was to put us in the condition of a thinker, a genius who spent her life without ever experiencing colors. We don't have to be strict on the quantity of information Mary gets, but on the quality, the topic she focuses on. This way the thought experiment becomes "a lot less far-fetched" (p.80). Mary may have missed the vector space information as well as many other physical facts, but this objection is nullified once we become more aware of the real intent of the argument. In addition to that, would she have known the feeling of looking at a poppy (I prefer this variation, already found in the literature⁷) via a third person approach? Our intuitive response seems to be negative and it doesn't have to do with the quantity of information we dispose of. If there's no experience, no contact with the object of interest, we can't give a phenomenological account that

5. It is common thought in the philosophy of mind that we may be in need of someone who will bring a new theory of consciousness that will revert our ideas on it, just like Newton did with gravitation.

6. See Chalmers (1996), Goff (2019) and Nida-Rumelin (2007).

7. For this beautiful example, see Dunne (1929).

is both coherent and justified. I could easily be accused of a strong empiricist approach, but would you ever say that an a priori concept of an apple is sufficient to account for your subjective conscious experience of it? Once again, it seems we need experience and another way to look at the knowledge argument.

Let's get straight to Section 3, where I consider a more intriguing line of objections, which found a lot of consensus among philosophers. It's time to analyze the phenomenal concept strategy.

Section 3. Third objection: the phenomenal concept strategy

The third and final objection I will consider in this paper is the so-called "phenomenal concept" strategy. The idea behind this line of thought is that Mary, once she is set free, comes to know something new, but her knowledge doesn't actually concern a new piece of information. In fact, the real progress she makes is granted by a new mode of presentation, a new way to learn a fact she already knew in her black and white room. This objection is also known as "New knowledge/old fact view"⁸. Before bringing in some examples from the literature, the idea can be stretched out with an example by Frege. We all know that Hesperus (the star of the evening) is Hesperus, that's a banal truism. After some empirical research, we are finally able to deduce that Hesperus isn't a different star from Phosphorus, the star we can see in the morning. They are the same and identical star; they are the same thing, no doubt about it. The question is: did we learn anything new? The answer seems to be negative, as we are referring to the same astral object in the sky, which is exactly the same both during the morning and the evening. We have simply come to know something we already knew (either by empirical observation or by a conversation with someone) without adding any new relevant fact. Sure, we know something we didn't know before, but that's not the point: our discovery

8. This is how the Stanford Encyclopedia of Philosophy reports it. It is commonly known as "Phenomenal concept strategy", but the idea expressed is the same.

doesn't imply any new fact, we are in a condition which is different, but with nothing new to offer and to analyze.

Now, are there any authors who embody this strategy? There are many of them. That's why I chose to analyze only some of these responses, as it would take too long to consider all the possible variations and the relative objections to each one.

Let's quickly structure the argument for the New Knowledge/Old fact view:

- 1) The phenomenal character (e.g. what is like to see red) is a physical property of experience.
- 2) In order to know what is like and acquire a knowledge of a phenomenal character, you have to acquire phenomenal concepts of that specific phenomenal character.
- 3) The acquisition of phenomenal concepts can be fully described in physical terms.
- 4) A subject has got a phenomenal concept only if he had the relevant experience of that very phenomenal character.
- 5) Mary gains knowledge of the phenomenal character of red under a phenomenal concept, which are both physical and describable in physical terms.

The conclusion has already been anticipated: Mary learns all there's to know about color and her discovery of the phenomenal character of red doesn't constitute a new fact, but it is an illusion. There's no space for anti-materialist claims, the epistemic progress consists of a new mode of presentation of the same old fact (learnt of course via chemistry, neurophysiology, science of vision etc...).

I will focus on three different approaches: Michael Tye (1995), Perry (2001) and Loar (1997). I can't cover all the other responses, as I mentioned before, but it's worth noting that the phenomenal concept strategy has become popular in the last ten years, becoming an answer more influent and accepted than, for example, Lewis and Dennett ones. Let's start with Tye's account.

Tye (1995) describes a phenomenal character as a capacity to represent internal and external objects, in a way "appropriately poised for use by the cognitive system" (see Tye 1995, 137–144). According to Tye, there are two kinds of phenomenal concepts: indexical (e.g. a specific yellow hue we have in mind) and predicative (our capacity to distinguish, for example, a hue from another one which is really close). Mary can't grasp phenomenal redness because she has never had the relevant experience. Although we may be tempted to accept the dualist conclusion, we have to stay alert and recognize another particular aspect of Tye's proposal, which can also be found in Van Gulick (2002) and Strawson (2020): experiences are concrete and physical. Their properties (in this case the phenomenal character associated) are physical and fully describable in physical terms, at least according to the defenders of the phenomenal concept strategy. In general, this is also the case for the physicalist resolutions on the mind-body problem. But Tye doesn't stop there: one could object that in order to understand what is like to see red, has got to be acquainted with that kind of experience. That's right: experience is needed, otherwise our comprehension remains silent and poor. How can we save physicalism? Here come phenomenal concepts: even though experience is necessary, it doesn't follow that qualia are non-physical. In order to know what is like, a subject needs the relevant phenomenal concept (e.g. what is like to see red is *this*), under which he can fully understand the physical nature of qualia. For a more revised account of Tye's ideas, see Tye (2009).

Perry (2001) takes inspiration by Tye's proposal and suggests that Mary's new knowledge can be accounted as a specific type of indexical concept, like "I am a philosopher" or "Today is sunny". Perry treats the new

information acquired by Mary as a case of demonstrative belief. After being released and being acquainted with the rose (which I find more poetic than the original ripe tomato, by the way), she's going to think "So having an experience of red feels like *this!*". "This", in the present context, refers to a physical property (the phenomenal character of the experience of red). It seems clear that Mary could have never known what it was like to have that chromatic experience before her release. What makes her thought true? The fact that experiences have that particular property. Therefore, says Perry, Mary doesn't learn anything new. The knowledge argument is valid, but its conclusion is denied once again. To be precise, the phenomenal concept strategy focuses on the second premise of the argument and on the step towards the third one. Mary gets familiar with a new mode of presentation, an indexical one: qualia are therefore physical.

Loar (1990 and 1997) proposes a complicated account for the knowledge argument, which can be summarized in the following concepts. In his mind, phenomenal concepts are *recognitional concepts*. The following passages will help us understand what he means by "recognitional":

They have the form 'x is one of that kind'; they are type-demonstratives. These type-demonstratives are grounded in dispositions to classify, by the way of perceptual discriminations, certain objects, events, situations. (p.225)

Recognitional concepts are generally formed against a further conceptual background. (...) A recognitional concept will then have the form 'physical thing of that (perceived) kind' of 'internal state of that kind' and so forth. (ibid.)

To have the phenomenal concept of redness is, in other words, to be able to recognize experiences of red while having them. For certain aspects it reminds of the account given by Lewis and Nemirow, which I

examined in Section 1. The main difference though is that in the proponents of the ability hypothesis account there's no mention whatsoever of phenomenal concepts, an innovation brought which is found in Loar writing. Since the philosopher predilates a physicalist account for qualia, he considers them as physical properties of experience. But what then? Loar claims that the recognitional concepts refer directly to their referent (the phenomenal character). This means that there is no other property involved in fixing the reference. The (physical) qualia triggers the phenomenal concept, which can then refer to that kind of property (the quale). Phenomenal concepts are, quoting Loar, "certain self-directed recognitional concepts" and "irreducibly demonstrative" (*ibid.*), like our high-order concepts (e.g. when I think about the concept "*phenomenal concept*"). If this account is correct, Mary's new knowledge simply consists of a bundle of phenomenal concepts, which aren't new facts for her immense knowledge of human vision. In the following part of this section, I consider some possible replies to Tye, Perry and Loar.

Replies: Chalmers (1996, 2002 and 2010) argues against the phenomenal concept strategy that it has got to face his framework concerning primary and secondary intensions. The framework Chalmers proposes states that an intension is said to be primary when it picks its referent in the actual world (see the case of H₂O in our world). The physical and the phenomenal concept would be distinguished in virtue of their primary intensions. Chalmers claims that if a single fact can be known under two modes of presentation (a physical and a phenomenal one), then we are confronting ourselves with two items of knowledge, which involve two completely different concepts. Because of the different primary intensions (I won't talk here of the secondary ones) we have two respective types of properties, the phenomenal (non-physical qualia) and the physical ones.

Another way to use Chalmers's framework can be found in Nida-Rumelin (2007), whose consequence is the refutation of Tye and colleagues' ideas. Note that a similar idea can also be found in Goff (2017). He claims

that the knowledge argument doesn't go against physicalism: but on the other hand, if we assume that phenomenal concepts are transparent, that is to say that they reveal the essence of the corresponding phenomenal property, then things turn out differently. Let's take a look at Goff's words:

in this case Mary's new knowledge is knowledge of the nature of red experiences, but if pure physicalism is true, she already knew the complete nature of red experiences in knowing the pure physical truths, and hence there ought to be nothing more she can learn about their nature.

(2017, p. 74–75)

A weaker response can be found in Alter (2013). Alter claims that the defenders of the knowledge argument can press on the role of experience. It's true that a phenomenal concept is formed once Mary experiences the qualia of red, but she has also got to have a first-person experience and get acquainted with the phenomenal character. Here seems that phenomenal concepts are insufficient to explain their epistemic progress (even though Alter's polemic target here are more precisely Tye and Ball (2009)).

Personally, I find the whole phenomenal concepts idea pretty well-structured. It doesn't refute the first premise like Dennett and Churchland surely would and it doesn't crush our intuition that Mary learns something new from her release. Overall, this is one of the most solid and powerful objections against Jackson's argument, but I must confess that I don't find it compelling. To begin with, a defender of this strategy should try to explain why we have the intuition that Mary learns something new, however materialist we may be (Jackson tries to answer in a postscript⁹). Second, there's a gap between what Mary knew in her room and when she has the relevant experience of red. And this gap can't be fulfilled with phenomenal concepts, as their nature consists of the qualia we experience. Mary, roughly, couldn't have formed the phenomenal concept of red when caged, because she lacked the relevant phenomenal character.

9. See the bibliography. In the postscript, Jackson looks for a possible explanation of our intuition on the knowledge argument, whose cause seems to be our misunderstanding of the nature of the object we experience.

If this very phenomenal character had been physical and tractable in objective terms, Mary shouldn't have the "aha!" response and a consequent epistemic progress. And she would have come to a strong physicalist conclusion way before the kidnappers decided to release her.

In sum, neither Tye, Perry nor Loar's objections seem to be a treat to the knowledge argument. The appeal to phenomenal concepts may sound strong and strike the most serious defenders Jackson's original intuition, but once we analyze the ideas behind it the whole castle collapses. The objection needs to explain our intuition on Mary's case, the nature of what is like and the sense of Mary's epistemic progress.

This brings us to the final section of the paper, in which I formulate an answer to the knowledge argument, using especially Gertler's acquaintance response. Can acquaintance be the answer we need to cast away the darkness and the confusion which surround the knowledge argument?

Section 4. The acquaintance response

The previous sections, if the answers to the various objections have proven successful, showed us that the ability hypothesis, the denial of Mary's epistemic progress and the phenomenal concept strategy should all be refuted. Is there a way to explain our intuition on the knowledge argument? Can we provide a coherent account for what is like to see red? The most promising approach to Mary's problem, to me, can be found in Gertler (2019). The idea is the following: since Mary can't deduce what is like to see red from her black and white room, we must assume that phenomenal knowledge needs something more. That extra ingredient is acquaintance, which corresponds to another kind of knowledge, both different from knowledge-that and knowledge-how (already seen in Section 1). In order to express acquaintance, a subject of experience has to be directly in contact with a phenomenal character (qualia). In Mary's case, she has to have in front of her

eyes a red object and she must also introspect that qualia, that particular feeling she has during the process. If she isn't acquainted with the famous ripe tomato, she can't even start to describe how it feels and what is like to have that particular chromatic experience. Quoting Gertler from Coleman (2019):

According to the acquaintance response, Mary's new knowledge is a genuine epistemic advance. Phenomenal knowledge- knowledge of what it's like- requires a grasp of the phenomenal property that can be achieved only through direct acquaintance with that property. It therefore requires that the target property is instantiated in one's own experience. This is why Mary gains phenomenal knowledge of redness only by experiencing redness herself; she cannot deduce this from the knowledge she has about redness before her release. (pp.64-65)

The response is also accepted by philosophers such as Chalmers (2003) and Goff (2015a). Robert Howell's subjective physicalism (see Howell 2013) also stipulates a similar conception, even though the conclusion is a monistic view of reality, with a strong rejection of any possible dualism.

The acquaintance response relies on an important idea exposed by Joseph Levine (1983), which is noted as the "explanatory gap". It seems that whenever we try to deduce conscious states and qualia from the physical facts, there's a gap between the two, like we were missing an explanation of how matter can be conscious. Science is impotent and silent on how the brain gives rise to specific conscious states (e.g. when I stare at a sunset and the wind is blowing on my face) and doesn't give an answer to the phenomenal aspect of our inner lives. This concept inspired Chalmers to formulate in a more precise manner (as in the literature the problem was already known by time¹⁰) the *hard problem of consciousness*, the problem of explaining why we are conscious instead of being machines without a subjective set of conscious states.

10. "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the djinn when Aladdin rubbed his lamp in the story."
— Thomas Henry Huxley (1825-1895)

Another strength by the acquaintance is the fact that it can respond to a powerful objection brought by Watkins (1989), which was partially responsible for Jackson's change of mind:

(...) If qualia are not causally efficacious, then my beliefs and memories would be just as they are whether there were qualia or not. (Watkins 1989, 160).

Since knowledge by acquaintance is achieved through a non-causal route (see Balog 2012 for more details), in acquaintance an instance of the property serves as a mode of presentation of that property. The relation between the thought about that property (qualia) and the property instance is then direct and non-causal. This simple fact enables the subject to directly grasp the phenomenal character of that experience, without necessity of any sort of causation at all (for a defense of this claim, see also Gertler 2012).

The acquaintance response rests on the idea that there's a strong ontological gap between the first and the third person perspective. Many philosophers such as Howell (2013) prefer the "epistemicism" solution, which corresponds to Chalmers' classification of Type-B materialists. Those who accept this idea also accept the first premise of the knowledge argument, but they don't see an ontological gap: instead, there's an epistemic one. Some facts of the world (the experiences in first person) can't be deduced from the physical facts, but this shows that a third person approach is inefficient. In order to grasp the blueness of the sky, you have to look at it and be acquainted with the respective (physical) quale. Physicalism is true and the conclusion of the knowledge argument is denied one more time.

I think there's a problem with this solution: why shouldn't we make a further step and assess an ontological gap? If physics is committed to the idea that an objective and impersonal view of the world is possible, then there shouldn't be any mention of subjectivity, in this very picture. But that's not the case. We know that without experience and without consciousness we couldn't even start to describe reality, as we wouldn't

even fell it in the way that we actually do. Consciousness and qualia are a fundamental feature of reality that are essential to both science and philosophy. If we can't neither reduce nor find an adapt physical explanation to qualia, we should embrace a dualist formula and appreciate the acquaintance response. I find that the response given by Howell should give a try to dualism, without fearing mental causation and an ontological economy prejudice which haunts physicalism nowadays. It's time to admit that consciousness and the mental are another part of our reality, which determines who we really are and who we shall be.

Conclusion

I may be too severe with physicalism and pretend that it doesn't give enough answers concerning our doubts on the knowledge argument and the real nature of qualia. On one hand, I would like physics and the metaphysical doctrine of physicalism to be a god-like instrument for human beings: perfect, which never commits mistakes and that will finally cast away our cartesian intuitions on the self and the mind. I wish it were so simple to refute the knowledge argument and its appeal, but I think it's impossible. Hempel's dilemma still sounds like a convincing argument to me and the hard problem posed by Chalmers seems to challenge even more the physicalist's account for consciousness. Then there is phenomenal knowledge, which is something more, a different part of reality which needs a subject that experiences it and, more importantly, of an explanation. Why does it feel like something to have an experience of a red rose? If we can't deduce this type of knowledge from any physical fact or any general quantity of physical information, then we must accept that consciousness is a fundamental feature of our reality, which establishes who we are (personal identity and subjectivity) and how we experience the world around ourselves. Without consciousness, qualia and the first-person perspective lose sense and our scientific understanding in third

person would also be meaningless. Our life would be spoiled and deprived of a unique aspect, a whole range of experiences and qualitative traces which allows us to feel alive, every day that passes by.

Sure, much of the debate on the knowledge argument is still going on and I haven't considered many other objections to Jackson's masterpiece. But if the previous sections convinced us to look for something more than physicalism, we are then ready to say that the knowledge argument is both valid and correct. Mary can't deduce what is like to see red from the physical facts: physicalism is false, and a new world of colors embraces our conscious mind.

References

- ✓ Alter T., 2008, "Phenomenal knowledge without experience", in E. Wright (ed.) *The Case for Qualia*, 247–67. Cambridge: MIT Press.
- ✓ Alter T. and Howell R., 2009, *A Dialogue on Consciousness*, Oxford University Press.
- ✓ Alter T., 2013, "Social Externalism and the Knowledge Argument", *Mind*, 122: 481–496.
- ✓ Alter T., 2017, "Physicalism and the Knowledge Argument".
- ✓ Ball D., 2009, "There are no Phenomenal Concepts", *Mind*, 118: 935–962.
- ✓ Balog K., 2012a, "Acquaintance and the Mind-Body Problem", in C. Hill & S. Gozzano (eds.), *New Perspectives on Type Identity: The Mental and the Physical*, Cambridge: Cambridge University Press.
- ✓ Chalmers D. J., 1996, *The Conscious Mind, in Search of a Fundamental Theory*, Oxford University Press.

- ✓ Chalmers D.J., 2002, “The Content and Epistemology of Phenomenal Belief”, in Smith, Q. & A. Jokic (eds.), *Consciousness: New Philosophical Essays*, Oxford: Oxford University Press, 220–272.
- ✓ Chalmers D.J., 2010, “Afterword to ‘The Two-Dimensional Argument Against Materialism’”, in D. Chalmers, *The Character of Consciousness*, Oxford: Oxford University Press, 192–206.
- ✓ Churchland Paul, 1989, “Knowing Qualia: A Reply to Jackson”, in P. Churchland, *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA: MIT Press, 67–76.
- ✓ Dennett D., 1991, *Consciousness Explained*, Boston: Little, Brown, & Co.
- ✓ Dennett D., “'Epiphenomenal' Qualia?”, In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by P. Ludlow, Y. Nagasawa, and D. Stoljar, 59-68. Cambridge, MA: MIT Press, 2004.
- ✓ Dennett D., 2007, “What RoboMary Knows”, in T. Alter & S. Walter (eds.) 2007: 15–31.
- ✓ Dunne J. W., 1929, *An Experiment With Time (second edition)*, London: A & C Black.
- ✓ Gertler B., 2012, “Renewed Acquaintance” in *Introspection and Consciousness*, edited by Declan Smithies and Daniel Stoljar. Oxford University Press, pp. 89-123.

- ✓ Gertler B., 2019, "Acquaintance, Parsimony and Epiphenomenalism", appears in "The Knowledge Argument", ed. Sam Coleman, Cambridge Classic Argument Series, Cambridge University Press, 2019, Ch. 3, pp. 62-86.
- ✓ Goff P., 2017, *Consciousness and Fundamental Reality*, Oxford: Oxford University Press.
- ✓ Goff P., 2019, *Galileo's Error: Foundations for a New Science of Consciousness*, Rider Books.
- ✓ Hempel Carl Gustav (1969), "Reduction: Ontological and linguistic facets", in S. Morgenbesser; P. Suppes; M. White (eds.), *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, New York: St. Martin's Press, pp. 179–199
- ✓ Howell R., 2013, *Consciousness and the Limits of Objectivity: The Case for Subjective Physicalism*, Oxford University Press.
- ✓ Jackson F. (1982), "Epiphenomenal Qualia", *The Philosophical Quarterly* Vol. 32 No. 127 (Apr.1982) pp 127-136
- ✓ Jackson F. (2003), "Mind and illusion", in *Royal Institute of Philosophy Supplement* September 2003
- ✓ Jackson F. (2007), "Colour for representationalists", in *Erkenntnis* (2007) 66:169–185

- ✓ Jackson F., “Postscript on Qualia”, In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by P. Ludlow, Y. Nagasawa, and D. Stoljar, 59-68. Cambridge, MA: MIT Press, 2004.
- ✓ Jackson F. (2019), “The knowledge argument meets representationalism about color experience”, appears in *The Knowledge Argument*, ed. Sam Coleman, Cambridge Classic Argument Series, Cambridge University Press, 2019, Ch. 5, pp 102-117.
- ✓ Lewis D., 1983, “Postscript to ‘Mad Pain and Martian Pain’”, in D. Lewis, *Philosophical Papers (Volume 1)*, Oxford: Oxford University Press, 130–132.
- ✓ Lewis D., 1988, “What Experience Teaches”, *Proceedings of the Russellian Society*, 13: 29–57; reprinted in W. G. Lycan, 1990b, 499–519, and in P. Ludlow, *et al*, 2004, 77–103.
- ✓ Levine J., 1983, “Materialism and qualia: the explanatory gap”, in *Pacific Philosophical Quarterly*, Volume 64, Issue 4, October 1983, pp. 354-361.
- ✓ Loar, B., 1990 [1997], “Phenomenal States”, *Phenomenal Perspectives 4: Action Theory and Philosophy of Mind* (1990), 81–108; revised version in N. Block, O. Flanagan & G. Güzeldere (eds.) *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press, 1997, 597–616.
- ✓ Nagel T., “What is like to be a bat?”, in *The Philosophical Review* vol. 83, No. 4 (Oct. 1974), pp. 443-450.

- ✓ Nemirow L., “Review of Nagel’s mortal questions” in *Philosophical Review* 89 (3) 473-7 1980.
- ✓ Nemirow L., “Physicalism and the cognitive role of acquaintance” in William G. Lycan (ed.), *Mind and cognition*, Blackwell 1990.
- ✓ Nemirow L. (2007), “So This is What it’s Like: A Defense of the Ability Hypothesis”, in T. Alter & S. Walter (eds.) 2007: 32–51
- ✓ Nida-Rümelin M., 2007, “Grasping Phenomenal Properties,” in T. Alter & S. Walter (eds.) 2007: 307–349.
- ✓ Perry, J., 2001, *Knowledge, Possibility, and Consciousness*, Cambridge, MA: MIT Press.
- ✓ Strawson G., 2020, “What does physical mean? A prolegomenon to panpsychism”, revised version of the chapter in *The Routledge Handbook of Panpsychism* (2020), forthcoming in *Mind and Being*].
- ✓ Tye M., 1995, *Ten Problems of Consciousness*, Cambridge, MA: MIT Press.
- ✓ Tye M., 2009, *Consciousness Revisited: Materialism without Phenomenal Concepts*, Cambridge, MA: MIT Press.
- ✓ Van Gulick R., in Peter Ludlow, Yujin Nagasawa & Daniel Stoljar (eds.), *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*. MIT Press (2004)
- ✓ Watkins M., 1989, “The Knowledge Argument against the Knowledge Argument”, *Analysis* 49: pp. 158-160.