OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Investigating the Mary Argument

Jackson Mackenzie Walstatter
terminator8734@gmail.com

ABSTRACT

This paper will tackle questions about the nature of experience and consciousness. More specifically, it will dive into the knowledge argument, as illustrated by the Mary argument. The Mary argument tries to establish the existence of subjective facts or qualia. Mary is a color scientist who knows all the physical facts about color. She is trapped in a black and white room. The question it draws: when Mary sees color for the first time, does she learn something new? There are two main responses to this argument: thick and thin materialism. Thick materialists posit that Mary will be surprised, while thin materialists say nothing will happen when Mary sees color. I will look at David Lewis's thick materialist argument that says Mary gains abilities rather than new facts. I will then look at Daniel Dennett's blue banana and locked RoboMary responses to the Mary argument that say Mary gains nothing new at all. My examination of Daniel Dennett's blue banana argument will show its inadequacy in refuting the knowledge argument. My examination of the locked RoboMary argument will show how it requires non physicality. A part of this examination will show that functionalism does not require physicalism, and that the functionalism in the locked RoboMary thought experiment requires something nonphysical. Lastly, I will demonstrate that the something that is nonphysical is a quality-less subjective experience.

INTRODUCTION

Physicalism holds that everything in the universe is physical or arises from the physical. Anything that seems nonphysical, a thought, for example, is just a physical state that we experience in the form of a thought. Non-physicalists deny that everything is or comes from the physical. In this paper, I will take the stance of a non-physicalist, looking at a past thought experiment on this topic. More specifically, I will tackle a physical interpretation of consciousness in reference to the Mary/Knowledge argument. The second section will go over the Mary/Knowledge argument against physicalism by Frank Jackson. The third section will go over three of Daniel Dennett's responses to the Mary argument against physicalism. These responses are the blue banana, RoboMary, and Locked RoboMary arguments. The locked RoboMary argument is a more airtight version of the RoboMary argument, so I will only look at the blue banana and locked RoboMary argument. I will show that the locked RoboMary argument requires nonphysicality. The fourth section will go over how the facts about the physical and functional states differ. I will show how the nonphysicality required in the locked RoboMary argument is something akin to a nonphysical consciousness.

Ⅱ. The Knowledge Argument

The battle between physicalists and non-physicalists plays out in different arenas. One of those arenas is known as the knowledge argument. Frank Jackson's knowledge argument will be the central focus of this paper, so we need to establish what exactly it is. The knowledge argument uses two thought experiments to prove its point. They follow two people: Fred and Mary. The Fred thought experiment demonstrates the knowledge argument quite well, but this essay will focus on Mary.

Mary is a scientist in quite a predicament. She's trapped in a room that's only black and white. The thing is, Mary's quite the brilliant scientist. While trapped in the room, she learns every piece of physical information, including all there is to know about color. She understands every single physical process that leads to color, from the light's travel to its arrival in the cornea, where the brain processes it, but all she has ever seen in her entire life was black and white. One day, her captors give her something new: a red rose. This leads to the question of whether, when she sees the red rose, she learns something new. Jackson says yes, Mary learns something new. His reason: she understands everything about why we see red, i.e., the physical mechanisms that cause the experience, but when she has the experience, she understands something new: what red *looks* like. Because she gains a new understanding (what red looks like) after having all the physical information, the physical information must leave something out. If physical information leaves something out, then physical information must be incomplete. If physical information is incomplete, it does not describe everything. If physical facts do not describe everything, then some other type of fact exists, and physicalism fails. (Jackson 1982, pg. 130)

The specific variety of nonphysical facts in this thought experiment are phenomenal facts. Thomas Nagel's *What It Is Like to Be a Bat* outlines them. In *What It Is Like to Be a Bat*, Nagel demonstrates how the physical reductionist account of consciousness fails to describe the phenomenal facts of an experience. To do this, he uses the example of a bat. His argument: we can understand everything about the mechanism of echolocation without ever knowing what it is like to use echolocation. Since we cannot understand the subjective situation of the bat using the physical facts, some type of information cannot be described by physical facts. This information he calls the subjective qualities of experience, what it is like to experience, or the phenomenal facts of an experience. (Nagel 1974, pgs. 438-440) This idea of facts or information describing the "what it is like" of an experience existing separately from physical descriptions serves as a foundational principle of Frank Jackson's knowledge argument. His argument calls these facts qualia.

I would like to take a moment to talk about the purpose of the knowledge argument. The knowledge argument attempts to disprove physicalism by demonstrating the existence of nonphysical facts. It does not try to establish what else it could be if it is not a type of information. It endeavors to show how nonphysical information exists. The thing to remember is that if you have all the physical facts and something lies out of your reach, then the physical facts are incomplete, and physicalism fails.

Two main camps emerged in opposition to the knowledge argument. Thick materialists deny that Mary learns anything new. They say that, regarding the experience of the rose, Mary learns nothing. Rather, she just feels surprised. A prominent thick materialist argument put forth by David Lewis states that Mary gains new abilities to represent the information she already knew, or she reaches a new representational state. I will touch on this argument more later, but for now I will say that this representational state surprises her because she hasn't experienced it before. Take, for example, John is playing Darth Vader's theme on the piano. John possesses all the physical facts about playing Darth Vader's theme on the piano. The act of playing the song represents the physical facts that create the audio. The thick materialist says that before playing the concert piano, John already knows how to play it, but when John plays it, he gets pleasantly surprised by the song. This raises an issue with the knowledge argument, like one we will discuss later. According to this interpretation, she already understands the facts about playing the piano, but when she plays, she represents that information in a way that surprises her. That representation serves as the experience, but the experience holds no new facts.

The thin materialists say that absolutely nothing will happen on release. Their argument for this conclusion is that she already has all the physical facts. Because she has all of the physical facts, and physical facts are all that describe experience, she can understand everything about the experience of color. Since she used the physical information to understand everything about the experience of color, the following must be true: 1. the physical information is complete in terms of color; and 2. absolutely nothing would happen upon release. Maybe she would shrug and say she expected this. To investigate this line of argument, we will use 3 of Daniel Dennett's arguments against Mary.

Ⅲ. Daniel Dennett's Blue Banana Argument

Daniel Dennett constructed two prominent arguments, with one being more recent and more relevant. However, both arguments merit discussion, so we will start with the older one. I call Dennett's first argument the blue banana argument. It starts with Mary's captors deciding to let Mary experience color, but they try to test her. They give Mary a banana, but it's not the regular yellow banana. This banana is actually blue. Mary's captors expect her to think the banana presented to her is just a regular banana, but she takes one look and recognizes its blueness. Her captors are baffled. How did she know? Simply put, Mary cross-checked all aspects of her physical state against her knowledge of all physical states. From there, she was able to say: My physical state is the same as the physical state of blue; therefore, I am not seeing yellow, but blue. (Dennett 1991, pg. 399-400).

Now this argument shows us the value of physical information, but it does not refute the claim that non-physical knowledge exists. Mary used her physical understanding to check and see what functional state she is in. The knowledge argument *doesn't* say she *won't* have a physical reaction to experience. It simply says that in the creation of experience, there is something *in addition to* the physical. Howard Robinson agrees with this idea. Robinson denies that Mary could piece together that she sees blue instead of yellow instantaneously. He argues that she cannot immediately recognize a color by applying the physical knowledge to her physical system. He admits that she could do it, but over an extensive period.

(Robinson 1993, pg. 175) Dennett wants us to conclude that the functional state of Mary's physical systems holds all the facts of her experience. In expression of this idea, Mary uses the physical facts to determine what it is like. Dennett tells us she checks in with her functional response to the situation. When she applies her understanding of the functional states she is in and the functional states of color, she then knows she sees blue. This leaves open the following argument: she knew the physical facts before the experience, upon the experience, she used the physical facts to deduce what she was seeing. So she needs the experience to deduce the qualia, therefore, she cannot know the qualia before the experience. As the contention lies in whether she knew the qualia before the experience, Dennett *did not* disprove the Mary argument. If he shows how she can know the qualia before or without the experience, then the knowledge argument has some issues.

Before we move on, I want to touch upon the idea that qualia give no new information. A prominent argument along these lines comes in the form of, "I think it is that knowing what it's like is not the possession of information at all. It isn't the elimination of any hitherto open possibilities. Rather, knowing what it's like is the possession of abilities: abilities to recognize, abilities to imagine, and abilities to predict one's behavior by means of imaginative experiments. (Someone who knows what it's like to taste Vegemite can easily and reliably predict whether he would eat a second helping of Vegemite ice cream.) "(Lewis 1983, pg. f131) In Lewis's view, the functional state is the experience, and the qualia are specific qualities of that brain state. For Vegemite, the experience is the functional state entered upon the interaction of Vegemite and the tongue. The qualia represent specific qualities of the experience, which are the exact qualities of the neural pathways. From here, he says that the qualia rewire the brain upon the experience of Vegemite, giving you nothing but the ability to recognize, remember, and imagine the experience upon the stimulation of these new neural connections. Since these qualities give us nothing but new abilities, we get no new facts from them, and with no new facts, we get no new information. But Lewis's argument leaves open a hole. Lewis does not establish whether the abilities acquired from the experience draw upon physical or phenomenal facts. We can easily say that the abilities of recognition, memorization, and imagination of red require phenomenal facts acquired from the experience of red. If you had never seen red, you could not recognize, memorize, or imagine it. This means you need the experience of red to acquire the abilities that come with it. Lewis's argument does not show what facts these abilities come from. She did not have the abilities before the experience. Because of this you could argue that upon the experience Mary gains the fact of what red looks like. After the acquisition of this fact, she gains the ability to recognize, remember, and imagine what red looks like.

There is something to this argument, but is not that qualia are not facts. Lewis's example shows that the information of an experience remains the same across experiences, the difference lies in how that information gets represented. Take for example the information describing the taste of vegemite. You can predict the taste of vegemite ice cream through recalling the information of what vegemite tastes like in its original form. But when you try vegemite for the first time, you gain a new fact, that fact being what vegemite tastes like. I think that facts are arrangements of information under the context of a certain situation. When tasting vegemite ice cream vs. vegemite, what does not change is the information about vegemite. What changes is the context and the situation of the information. What vegemite tastes like is

the same whether in ice cream or original form. The context and situation of vegemite ice cream and vegemite do differ. Now if there is a difference between any one of the information, context, and situation of two facts, then those two facts *are* different. This means that the fact of what vegemite tastes like in original form and the fact of what vegemite tastes like in ice cream form differs. Since these facts do differ, when you taste vegemite ice cream, no matter if you already tried vegemite in its original form, you now know what vegemite tastes like. Now if Lewis shows how you can know what vegemite ice cream tastes like before you try vegemite ice cream, then he has refuted the knowledge argument, but alas he does not. So, after looking at Dennett and Lewis's viewpoints, we now see that the way for physicalism to undermine the knowledge argument is by showing that Mary can know everything about the experience, and the qualia of it, without having the experience and the qualia of it.

Ⅳ. Locked RoboMary Argument

Dennett's other main argument against Mary is called the RoboMary argument. We see two versions of this argument, locked and unlocked RoboMary. Locked RoboMary is a stronger version of the unlocked RoboMary argument, so we do not need to look at the unlocked RoboMary argument. RoboMary is a Mark-19 robot. All Mark-19 robots have color vision, but RoboMary cannot see color. She cannot see color because of a system that restricts her color cameras to a grey scale. If you only see grey, the only functional state of visual perception you can enter is grey. A good real-world metaphor for RoboMary's perceptual predicament is colorblindness. Now, this functional perceptual state of grey Dennett calls State A. Now RoboMary really wants to see color. To see color, she sets up a model of herself on some spare RAM. This model models every single functional state of the perceptual experience of red. This modeled perceptual experience of red is called State B. Basically, RoboMary is physically locked in a way that she can *only* see grey, so she constructs a model of herself that models the physical predicament of seeing red.

To understand how RoboMary can enter state B herself and know what it is like to see red from state B, we will use Dennett's example of RoboMary's interaction with a tomato. RoboMary sets down a tomato on a table. She instantly goes into a cascading series of physical states that then create the functional state of grey. These cascading physical states are state A. At the same time, RoboMary has the red seeing model of herself running. In opposition to state A, given RoboMary understands the tomato should be red, she runs her model, entering all the cascading physical states in a regular Mark-19, creating the functional state of red. Because RoboMary has access to all the physical facts, she understands every single physical difference between state A and state B. The knowledge of every physical difference allows her to then correct those differences, and put herself into state B or the functional state of seeing red. Now that she is in the functional state of red, Dennett says that she knows what it is like to see red. (Dennett 2005, 127-129)

To understand what Dennett wants to do here, let's look more closely at his position. Dennett is a physically oriented functional illusionist. Functionalism is the idea that every experience we have is just the function that it serves in the brain. For example when you see red, you are now in the first functional

state of red color perception, so the functional state *is* the experience. To clear up ambiguity for the reader, I will mesh these two terms, calling it the functional experiential state.

Now, Illusionism is the idea that subjective experience is a generated illusion of a complex physical reaction. Dennett's physical orientation of these viewpoints says that the functional illusion comes from *only* the brain state. The knowledge argument says that qualia and facts of them are nonphysical. The locked RoboMary paper is a response to this argument, so it needs to refute the notion of nonphysical qualia.

So, in order for Dennett to refute nonphysical qualia, he *must* demonstrate that qualia come from only the state the brain is in. Furthermore, Dennett cannot refute the existence of qualia, as no matter if they are an illusion or not, the illusion exists, so the qualia exist. What he can do is show what fundamentally creates qualia. If Dennett demonstrates that Qualia fundamentally come from the brain state alone, *then* we can accept physicalism.

So, Dennett must show that qualia come from the brain state alone. In the thought experiment, RoboMary enters state B and knows what it is like to see red. If this is true, then the qualia are physical, as Dennett's viewpoint requires all functional states to be physical. If the qualia are physical, the issue they present is solved.

Ⅴ. Diving Deeper Into RoboMary

The foundational principle of Dennett's argument and conclusions:

- The functional state comes from the brain state

From the thought experiment, Dennett wants us to draw the following conclusions:

- RoboMary can know the qualia of red by entering the cascading functional state of red
- RoboMary knows the qualia of red through the physical facts of the functional state
- One can know the qualia from knowing the physical facts
- Qualia are a type of physical facts
- Since qualia are physical, Frank Jackson's interpretation of nonphysicality is wrong.

Dennett's argument here is quite convincing. It solves the issue of facts about qualia by making facts about qualia a type of physical fact. But for RoboMary to know everything about qualia from physical facts, all facts about an experience must come from the brain or physical state alone. Using a model of color perception:

- A photon of 720nm hits the retina, stimulating photoreceptors within it.

- The retina transforms the stimulation into a neural signal containing all the information about this stimulation.
- Neural pathways send the neural signal to the brain.
- Neural circuits in the brain get activated and process the information contained within the signal. This step produces the complete physical state.
- The physical state produced in the previous steps produces the specific functional state of red.
- Upon the functional state the perceptual experience of red occurs.

In this model, RoboMary's locking mechanism kicks in at step four. This is for two main reasons. The first is that state A is the functional experiential state of grey color perception, while state B is the functional experiential state of red color perception. So the main difference between these states is not the physical brain state, rather the functional experiential state. The second reason is that for locked RoboMary to have the possibility of entering state B, she must be able to enter the physical brain state of red color perception. Dennett says that the physical brain state is the functional experiential state, so if she can enter the functional experiential state, she *cannot* be locked from entering the physical brain state. If she were locked from entering the physical brain state of the functional experiential state of red, from a physicalist perspective, the switch from state A to state B would be impossible.

To cement that RoboMary is locked at step four, we will look at locked RoboMary's predicament if she is locked at any of the previous steps. Now if the locking mechanism kicks in at any step before step four, then state A does not have complete access to the physical brain state of red color perception. If the locking mechanism occurs at any of these earlier steps, for her to then reach state B, one of two things must happen: 1. She unlocks her physical brain state and enters the functional state. Since she removed the locking mechanism itself, she then enters the perceptual state of red, or 2. She does not unlock herself, staying physically restricted from entering the functional experiential state, but she enters the functional state anyway by putting herself into state B.

Number one cannot be the case because Dennett says that she does not enter the perceptual state of a red color experience.

So, if she is locked from steps one through three, number two must be the case. In number two she is locked from entering the physical brain state of the functional experiential state of red. But she still enters the functional experiential state of red. Scenario number two says she can enter a functional experiential state, her brain cannot physically be in. By a physicalist definition this is *impossible*. In order for scenario two to work for a physicist definition, she must unlock herself. But if she unlocks herself she then enters the perceptual experience that Dennett outright says that she does not enter.

Either way you look at it, by what Dennett says occurs in the thought experiment, she cannot be locked at any step before step four.

So the locking mechanism occurs when locked RoboMary enters the physical brain state of the functional experiential state of red. So she has access to the complete physical brain state without access to the functional experiential state. Here we see the main thing that Dennett overlooks in his haste to refute the Mary argument:

The locked RoboMary thought experiment requires a difference between the physical brain state and the functional experiential state, that is *not physical*. In other terms, for Dennett's thought experiment to work *something* (I cannot say what it is *yet)* nonphysical exists. If something nonphysical did not exist, then the moment RoboMary enters the physical brain state she enters the functional experiential state, but Dennett requires RoboMary *not to*. So even if the locked Robomary thought experiment refutes the knowledge argument, *something* nonphysical exists.

Now we must face the question of: Does Daniel Dennett refute the knowledge argument by showing that qualia come from the physical alone?

Simply put he does not. Since we have just shown that the functional experiential state requires a nonphysical difference in comparison to the physical brain state, facts about the functional experiential state *are not* physical. We need to backtrack here. The functional experiential state is generally treated as an outcome of a process of its creation. But we have just shown the two main components of its creation, the physical brain state, and whatever the nonphysical component of the functional experiential state is. Since the functional experiential state *is not just* the physical brain state, the functional experiential state *cannot* be described in *strictly* physical facts, there must be something nonphysical about these facts. The nonphysical aspect of the facts about the functional experiential state are the qualia.

VI. The Computer Metaphor

To show how the nonphysical aspect of facts about the functional experiential state are the qualia I will use the metaphor of a computer. Four things will get represented in this metaphor:

1. The functional experiential state,
2. The physical Brain state,
3. The nonphysical component, and
4. Qualia.

Looking at a computer display we see two main things that are required to create it. The first is the code or the physical predicament of the computer. The second is electrical energy. Only after these two components interact do we see something pop up on the display.

Now something to note is that the computer *can* be in the physical predicament of google docs, without google docs appearing on the display. Google docs only appears on the display *after* the electrical energy

has run the physical predicament of the computer. This means that you have access to facts about using google docs *after* a process including the electrical energy occurs.

When we translate the analogy of the computer to the creation of experience we see that the functional experiential state is the display, the physical brain state is the code or physical predicament of the computer, the electrical energy is the nonphysical component, and the qualia are the facts about using google docs.

We see that the functional experiential state *only happens after* the interaction of the physical brain state and the nonphysical component. This means that facts about the functional experiential state *are not just* the facts about the physical brain state. Qualia are facts about the functional experiential state. Since qualia are facts about the functional experiential state, and facts about the functional experiential state *are not just* the facts about the physical brain state, RoboMary *cannot* know the qualia from just the facts about the physical brain state. This means that Dennett *fails* at proving qualia are physical, and the knowledge argument holds.

Ⅶ. Implications of Locked RoboMary

But within the locked RoboMary argument we see bigger implications for the theory of functionalism. In order for the functional experiential state to be the experience itself, the functional experiential state *must contain some* nonphysical component. This means that the theories of functionalism and physicalism are not compatible, you must choose one or the other.

We are still left with a looming question: since qualia are outcomes of nonphysicality, what is the nonphysical component? To answer this question I will revisit the computer metaphor. When looking back at the computer metaphor we see that you can be in the physical brain state of a functional experience, without being in the functional experience. You enter the functional experience *after* the nonphysical component gets introduced.

In western philosophy the nonphysical component is often associated with subjective experience. In the knowledge argument we see the perspective that qualia *are* the nonphysical component. But we just demonstrated that the qualia are *outcomes* of the interaction between the nonphysical component and physical brain state. Qualia are the qualities of subjective experience.

One of the most prevalent arguments against nonphysicality in Western philosophy is that the qualia's qualities are physical. I agree with this argument, when looking at the computer metaphor, we find that all of the nature of the qualities of google docs exist within the code of the computer. The font georgia exists within the code, but it needs the nonphysical component to appear on the display.

Since qualia are qualities of the subjective experience an equation of qualia would look like this: qualia = the qualities (originating from the physical brain state) + subjective experience. Now a major issue with

physicalism is the hard problem of consciousness (Chalmers 1995,pg. 201) . The hard problem asks: if everything is physical, why do we have subjective experiences?

This is the one thing physicalism cannot account for: subjective experience itself. We have now seen this in our "equation" for qualia and the hard problem. Since something nonphysical *must* exist, and the one thing physicalism cannot account for is subjective experience, in its purest form, the nonphysical component *is* a pure subjective experience that holds *no qualities*.

Our revised equation for qualia now looks like: qualia = qualities (coming from the physical brain state) + a quality-less subjective experience.

Something to note is that you can have the physical brain state without the functional experiential state, but if you have quality-less subjective experience then you have the functional experiential state. This means that in the creation of the functional experiential state, the quality-less, subjective experience is more fundamental than the physical brain state.

CONCLUSION

This paper has tied together the implications of the locked RoboMary thought experiment into the knowledge argument. It explores an implication of the locked RoboMary argument Dennett overlooked. The implications expounded on in this paper point us towards a nonphysical interpretation of consciousness. The most important thing to conclude from this paper is that a physicalist perspective accounts for the qualities of a subjective experience, but not subjective experience itself. Considering that the reason subjective experience exists is *not physical* we may conclude that subjective experience is nonphysical.

REFERENCES

Chalmers, David. 1 facing up to the problem of consciousness David J. Chalmers, 1995. https://consc.net/papers/facing.pdf.

Dennett, Daniel. ConsciousnessExplained.pdf, 1991. https://andrewmbailey.com/papers/ConsciousnessExplained.pdf.

Dennett, Daniel. What robomary knows | Sweet Dreamsphilosophical obstacles to a science of Consciousness | Books Gateway | MIT Press, 2005. https://direct.mit.edu/books/monograph/1947/chapter/53374/What-RoboMary-Knows.

Jackson, Frank. Epiphenomenal qualia Frank Jackson the philosophical ..., 1982. https://www.sfu.ca/~jillmc/JacksonfromJStore.pdf.

Lewis, David. Mad pain and martian pain | philosophical papers volume i | oxford academic, 1983. https://academic.oup.com/book/36015/chapter-abstract/313260255?redirectedFrom=fulltext.

Nagel, Thomas. Philosophical review what is it like to be a bat? author(s): Thomas Nagel, 1974. https://philosophy.uconn.edu/wp-content/uploads/sites/365/2020/03/Nagel-What-is-it-like-to-be-a-bat.pdf.

Robinson, Howard. Dennett on the knowledge argument | analysis | Oxford academic, 1993. https://academic.oup.com/analysis/article-abstract/53/3/174/95328?redirectedFrom=PDF.