Francisca Vasconcelos

Minds & Machines

Recitation 1

4/21/2017

Jackson's Qualia

Before trying to determine if machines can be "intelligent," we must understand what exactly makes people intelligent. Many believe intelligence can be attributed to the idea of consciousness, but what is consciousness and where does it come from? In this paper, I will explore the topic by 1) defining consciousness, 2) explaining the physicalist view, 3) describing Frank Jackson's view of consciousness, and 4) explaining why Jackson's view is flawed, preventing the rejection of physicalism.

Before diving into the different theories of how consciousness arises, we must first understand how philosophers define consciousness. While the term "consciousness" has many meanings in ordinary language, in what follows we will use the definition of *phenomenal consciousness*. As described by Thomas Nagel, "an organism has conscious mental states if and only if there is something that is like to be that organism–something it is like for the organism" [1, pg. 436]. From this arises the hard problem of consciousness: why is there "something it is like" for a subject in conscious experience? Since this question is far from resolvable with the current understanding of neuroscience, philosophers have developed various schools of thought, including physicalism and epiphenomenalism, to explain consciousness.

We will begin by exploring the physicalist view. This holds that "consciousness can be explained by the standard methods of neuroscience and psychology" [31, Chalmers Experience] (once the science is advanced enough), because "sensations and other mental states are entirely physical

[or]...every type of mental state is identical to some type of physical state" [2, pg. 303]. To illustrate the idea, we will look at the commonly used example of pain. A physicalist (specifically an identity theorist) would claim that pain is equivalent to the stimulation of C-fibers (nerve cells). When you cut your foot or touch a hot oven, causing a lesion to your skin, the C-fibers neighboring the lesion send signals to the nerve cells, causing the sensation of "pain," a mental state. That is to say, "pain just is–is nothing over and above–C-fiber stimulation...mean[ing] that pain couldn't possibly be present in the absence of C-fiber stimulation, or vice versa" [2, pg. 304]. Physicalism is often referred to as a reductive theory because it claims that phenomenally conscious states supervene on physical states. This implies that there could be no p-zombie replica of our world and that a molecule-by-molecule replication of any system will have exactly the same physical and conscious attributes.

Frank Jackson, on the other hand, belongs to the dualist school of thought. He believes that "there are certain features of bodily sensations especially, but also of certain perceptual experiences, which no amount of purely physical information includes" [127, Jackson]. These physically inexplicable sensations (i.e. smelling a rose, hearing a melody, tasting chocolates) are what he calls *qualia*. He strongly opposes the physicalist philosophy and in his "Knowledge Argument" defends the idea of qualia.

In particular, Jackson introduces the famous "Scientist Mary" thought experiment, which goes as follows: Imagine there is a scientist named Mary, who knows every physical detail there is to know about vision. She can tell exactly how the red wavelength of an apple will stimulate the cones of the retina, causing signals to be sent to the visual processing region of the brain, allowing the viewer to understand it is an apple. However, Mary has spent all her life living in a black-and-white world. Thus, she knows everything there is to know *about* seeing color, but she has never actually *experienced* color. So, if Mary is let into the real, color-filled world does she

learn anything new? According to Jackson, "it seems just obvious that she will learn something new about the world and our visual experience of it" [3, pg. 130]. And because Mary had all the physical information beforehand, learning something new implies that there is more than just physical information, meaning physicalism is false. This thought experiment can be summarized in a concise premise-conclusion form:

Premise 1. *Mary knows all physical facts pertaining to vision.*

Premise 2. *According to physicalism, physical facts are the only facts.*

Conclusion 1. *According to physicalism, Mary knows all facts pertaining to vision.*

Premise 3. *Mary learns something new when seeing color for the first time.*

Premise 4. *In learning something new, Mary learns a new fact.*

Conclusion 2. *Mary learns a new fact, meaning she did not know all facts pertaining to vision.*

Conclusion 3. *Physical facts are not the only facts.*

Conclusion 4. *Physicalism is false.*

While Jackson's believes that this argument for phenomenal consciousness poses a threat to physicalism, there are holes in the thought experiment that prevent the rejection of physicalism.

The issue I take is that Jackson's "Knowledge Argument" is based on two contradictory premises. If **Premise 2** holds then **Premise 1** requires that Mary knows all the physical facts, *as defined by physicalism* [**Conclusion 1**]. There is, however, no reason to believe that this holds in the "Scientist Mary" thought experiment. On the contrary, **Premise 1** is contradicted by what is currently known about the visual system. To understand how this contradiction arises, we must note some scientifically demonstrated facts of the visual system. Most notably, neuroscientists have long identified regions of the brain that respond *only* to color. It is believed that these are the

regions where color perception occurs. Even if you doubt this assertion, let us assume, for the sake of argument, that it holds. We will call this assertion the "existence of color-only brain circuits" or the "color circuit assertion," for short.

Now that we are familiar with the "color circuit assertion," we can pinpoint some of Jackson's flawed assumptions of the physicalist view. Since Mary has lived all her life in a black and white room, she has never stimulated the "color circuits" of her brain. This does not prevent her from knowing what Jackson calls "all the physical facts pertaining to vision." She can know about the existence of the "color-processing circuits" and she can even understand their inner-workings in infinite detail. However, because all these facts were acquired from black-and-white books, she has *never stimulated* her own "color-processing circuits." Consider now the physicalist definition of physical facts. For this, we will simply transcribe the pain example ([2, pg. 304]) mentioned earlier. "A physicalist would claim that color perception is equivalent to the stimulation of the color circuits. When you see color, the circuits send signals to the remaining brain, causing the sensation of 'seeing color,' a mental state." That is to say, "color perception just is–is nothing over and above–color circuit stimulation, meaning that color perception couldn't possibly be present in the absence of color circuit stimulation, or vice versa." Under the "color circuit assertion" it is very clear that Mary cannot know all the physical facts, in the sense they are defined by a physicalist, until she leaves the room and experiences color.

In summary, under the "color circuit assertion," the problem is that Jackson's definition of "knowing all physical facts" does not match the physicalists' definition. This is because it is possible to understand the inner-workings of color processing in infinite detail without ever using the "color circuits." It suffices that the brain uses different regions to 1) understand color processing and 2) implement color processing. Since it is possible to learn how color processing works from black-and-white books, and this is indeed Jackson's main assumption, it can be concluded that

understanding all the details of "color processing circuits" does not require using the said circuits. Hence, although Mary knows all the physical facts about color under Jackson's definition, she does not know all the physical facts about color under the physicalist definition.

On the contrary, the "something that Mary learns" when exposed to color is just the stimulation of the "color circuits" (what physicalists would consider the key physical fact for color perception). Mary can now correlate her understanding of color perception with the experience of seeing color (a purely physical process). When viewing the color red, Mary learns nothing more about how her eyes and brain are processing the red photon wavelength. However, her eyes and brain are finally undergoing the series of bio-chemical reactions described by the process, allowing Mary to actually experience the sensations she had read about. In this way, we have proven that, under the "color circuit assertion" Jackson's **Premise 1** does not hold for the "Scientist Mary" experiment. It follows that, there is at least a sensible assertion (that color is perceived with resort to "color circuits"), fully supported by modern neuroscience, under which the "Scientist Mary" experiment is fatally flawed.

To further clarify this point, let us explore the case study of swimming. Imagine a boy named Markus who has lived all his life in the desert, with minimal access to water (solely for the purpose of drinking). However, he has read every book and watched every YouTube video pertaining to the topic of swimming. Thus, Markus knows everything there is to know about swimming. For example, he can describe how every muscle in the body must move to perform a breaststroke and exactly how the body should be positioned to stay afloat when doing freestyle. One day, Markus goes to a swimming pool and attempts to swim for the first time. He immediately drowns. However, over the course of a week Markus practices the form he already "knew" and manages to swim back and forth across the pool.

What has happened? The problem is that, although knowing what Jackson calls "all the

physical facts" about swimming, Markus has never stimulated the regions of the motor cortex needed for swimming before entering the pool. By jumping in the pool, he is forced to use these undeveloped regions of his brain and cannot find a way to translate his knowledge of swimming into the physical act of swimming. With practice, though, he develops the muscles and synapse connections needed to swim. Under the physicalist view, these are the physical facts that matter for swimming. The continuous stimulation of muscles and motor neurons enables Markus to become a swimmer. Markus solely gained the physical ability to carry out the facts he already understood. The only information gained is the repeated physical stimulation of the motor cortex. This supports the physicalist view.

In this paper, I introduced phenomenal consciousness, the physicalist view of consciousness, Frank Jackson's view of consciousness (particularly the "Scientist Mary" thought experiment), and a swimming case study, in order to demonstrate the flaws of Jackson's reasoning, preventing the rejection of physicalism. While neuroscience may presently be too primitive to explain the problem of consciousness, I believe that it will eventually be able to, as the field matures. Before scientists knew the detailed biological structure of the eye and visual cortex, they must have found it absurd that people could process photons into the images we perceive mentally. They could have referred to this mysterious process as "visioness" and tried to explain it in non-physical terms. However, with modern science, we know this process is simply a complex chain of chemical reactions and electrical pulses in the brain. Although consciousness is still an open problem with no hard scientific results to support any claims, in the long-run I have faith that it will be explained by the *physical* properties of the brain. This implies that there is still hope for conscious and truly intelligent AI.

References

1. T. Nagel. What Is It Like to Be a Bat? *The Philosophical Review*, 83(4):435–450, Oct 1974.

2. B. Gertler. In Defence of Mind-Body Dualism. *Reason and Responsibilty*, 13:303–315.

3. F. Jackson. Epiphenomenal Qualia. *The Philosophical Quarterly*, 32(127):127–136, Apr 1982.