

If we accept that Mary, the color scientist, gains new knowledge when she sees the color red for the first time, must this lead us to a non-physicalist theory of consciousness?

Sam Wilkinson

*University of Edinburgh*¹

A common and popular option in defending Physicalism against the Knowledge Argument (KA) is the “phenomenal concept strategy” (PCS). PCS claims that, although ex hypothesi Mary knows all the propositions pertaining to color and experiences of color, there is at least space for the claim that she acquires a new concept, and thereby accesses these propositions under different, phenomenal modes of presentation. In short, Mary acquires new concepts upon her release and that explains her “discovery.”

Here I will show there is a way of saving Physicalism that does not appeal to PCS in the standard sense but entails that Mary acquires the ability to think a new (and philosophically under-appreciated) kind of singular thought. In acquiring this, she gains a kind of indexical, egocentric knowledge.

One prominent physicalist has recently rejected his earlier acceptance of PCS. Michael Tye (2009) claims that Mary already has phenomenal concepts deferentially. He says that “Maybe *fully* understanding a general phenomenal concept requires having had the relevant experience; but if such concepts are like most other concepts, possessing them does not require *full* understanding. They can be partially understood” (p.63). Instead of PCS, Tye puts forward the notion of acquaintance, a non-conceptual epistemic relation that we have to things with which we come into contact. He says that the KA “assumes that all worldly knowledge is knowledge *that*. I now think that this is where the knowledge argument crucially goes wrong” (2009, p.131). What Mary comes to know is (for example) the property red: She comes into perceptual contact with the property red, as typically instantiated in red things. She does not gain any new concepts: She just increases her understanding of concepts that she already possesses. Call this the *acquaintance strategy*.

Central to Tye’s rejection of PCS is his acceptance of *radical transparency*. What is meant by transparency is as follows: When one attends to one’s conscious perceptual experience, one is aware of features of the world perceived. Radical transparency claims that one is only aware of such worldly features. We can take it to be a rejection of something like Chalmers’ claim that “phenomenal redness (a property of experiences, or of subjects of experiences) is a different property from external redness (a property of external objects), *but both are respectable properties in their own right* [italics added]” (Chalmers, 2002). The claim of this final clause is the myth that Tye and other transparency theorists are keen to dispel. Our

¹ s.l.j.wilkinson@sms.ed.ac.uk

experiences of red things are (pleonastically) qualitatively red, but there is no intrinsic qualitative property of my experience that is phenomenal red. This is also very naturalistic: There is the organism and there is the world. The organism perceives the world and interacts with it to a greater or lesser degree of correctness or adaptivity: It can, therefore, be said to represent worldly objects and properties. The experience, qua brain state, has a number of properties: physical, relational, representational, temporal (for example, the property of having obtained five minutes ago). They are properties of the vehicle and in no way resemble, nor are in any way isomorphic (however this is conceived) with, the worldly properties represented. I will accept radical transparency without further argument.

Although I believe that the acquaintance strategy is a step in the right direction, what is not strictly ruled out is that Mary subsequently conceptualizes her experiences; namely, that she forms thoughts about the elements of the world that she has newly encountered in a way that doesn't make use of the general color concepts she (according to Tye) possesses deferentially. Tye has not taken things far enough: The KA does indeed rest on the assumption that all knowledge is knowledge that, but this hides a deeper problem, namely, an allocentric bias regarding the enabling conditions for certain thoughts. Many (indeed most) of our thoughts are singular, acquaintance-based thoughts with contextual enabling conditions. In short, when Mary comes out of the Jackson room and perceives colors for the first time, she necessarily acquires the ability to think new thoughts about these colors.

I will now sketch a theoretical framework that supports this view and test it out on a thought experiment that Tye's framework can't account for.

Modes of presentation (concepts, if you will) are thought constituents and can be understood as mental files. I open a mental file for an individual upon (direct or indirect) acquaintance with that individual, and I fill that file with information: It is the file that fixes the referent, not the information in the file. The inferential roles are altered while the file keeps its identity. Somebody who doesn't know that Hesperus and Phosphorus are one and the same celestial body possesses two different mental files: that explains why it is not irrational for someone, who doesn't know that they happen to co-refer, to assert something of "Hesperus," while denying it of "Phosphorus." On the other hand, I take my thought about John at t_1 to still be about John at t_2 , no matter what I have learnt about him in the meantime. As Campbell (1987) says, I am simply "trading on co-reference," and this is a condition for the possibility of singular thought: To think of something, you have to mentally track it.

The mental files metaphor is commonly used for individuals but it can also work for kinds and properties. As one learns more about "elms," that information goes into the "elm" file; as one learns more about what people acceptably call "red," that information goes into the "red" file. But that is not going to be of help here. Mary already has encyclopedic knowledge of red surfaces and experiences of red surfaces: Both her "red" file and "red experiences" file are ex hypothesi fully stocked. Her new thought is not going to involve that sort of mental file. My claim is that the new thoughts that Mary can form are a special kind of singular thought, using a special kind of acquaintance-based file. I will call these (following Recanati,

in press) egocentric files. An egocentric file is a repository for information gained from perception, from acquaintance with the world. It is implicitly situated (and self-locating) owing to an obvious but overlooked point: We are lumps of matter perceiving the world, and we can only be in one place at a time. In order to have a thought with one of these egocentric constituents, one must be standing in a suitable perceptual relation (direct or indirect) to the referent.¹ Now, with the case of Mary's new concepts, what are these referents? If we accept radical transparency, they cannot be phenomenal properties of experiences. They must, therefore, simply be properties represented in experience; namely, they are properties of things. Mary comes out of her room and for the first time stands in an epistemically rewarding relation (of the right kind) to red things. New thoughts are thus enabled. But these egocentric files are short-lived: They are based on a certain acquaintance relation and their role is to store information acquired in virtue of that relation. They, therefore, only exist as long as the relation holds.² While the relation does hold, however, these can be merged with a more stable encyclopedic concept to form a recognitional concept. Tye claims that Mary already possesses this encyclopedic concept deferentially. We can grant him this: Nonetheless, the egocentric file is certainly a precursor to the recognitional concept.

If you are not convinced that this is the case, consider the following variation on the Mary example taken from Nida-Rumelin (1996). Instead of coming out of the black-and-white room into the outside world, she goes into another room with colored patches on the walls. Surely you will not deny that she can form many thoughts about these colors (these patches) with which she becomes acquainted. She may notice that the pink and red patches resemble each other more than, say, the orange and blue. She may form preferences about experiencing certain shades. She does not, however, know which shade is which: what they are called, what their wavelengths are. She will only know what red is and that "that color" is (called) "red" when she sees ripe tomatoes, fire engines, London buses, that is, recognizably "red" things.

When Mary comes out of the black-and-white room for the first time and is confronted with a London bus, she says to herself (e.g. deduced from her knowledge that "London buses are red"): "So *this* is RED." What happens here is that she opens a new egocentric file that could be expressed as "This color." In this particular instance, the file then gets very swiftly "merged" with her encyclopedic file (packed with scientific information such as "light of 630-740 nm wavelength hitting a normal retina, etc...") and with deferential information such as "what these guys experience as, and call, 'red' "). The result is that she forms a new recognitional concept, "*So this is what these guys experience as red and what light of 630-740 nm wavelength hitting my retina does to me.*"

References

- Campbell, J. (1987). "Is sense transparent?" *Proceedings of the Aristotelian Society*, 88, 2-21.
- Chalmers, D. (2002). The components of content. In D. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings*. Oxford: Oxford University Press.

- Nida-Rumelin, M. (1996). What Mary couldn't know. In T. Metzinger (Ed.), *Phenomenal consciousness*. Paderborn, Germany: Schoenigh.
- Recanati, F. (in press). In R. Jeshion (Ed.), *New essays on singular thought*. Oxford: Oxford University Press.
- Tye, M. (2009). *Consciousness revisited: Materialism without phenomenal concepts*. Cambridge, MA: MIT Press.

¹ Egocentric files work for many indexicals: “Here” relies on being and perceiving the place where one is. “There” relies on perceiving (often very indirectly) the place denoted. “This man” relies on perceiving the man demonstrated, and so forth.

² Note that I can establish such a relation by iconically imagining a determinate shade of blue. This is simply an acquaintance relation with an uninstantiated property.