

# Curiosity and the Knowledge Argument

It was the night before her release from the black and white prison where Mary had spent her whole life thus far. Tomorrow the door would open, and the coloured light would flood in. Mary was excited; she was nervous. But most of all, Mary was *curious*. What would it be like to see red, or green or blue? Mary of course knew a lot about the quantitative structure of colour experience, e.g., how much hue, saturation and brightness each colour had, relative to other colours. And she'd sometimes played a little of game of trying to imagine colour by 'filling out' that structure with the qualities Mary was acquainted with from her sound experience: replacing in her imagination brightness with loudness, hue with pitch, etc. But she knew in her heart of hearts that this abstract exercise could never really provide her with insight into the character of the qualities that fill colour experience.

The morning finally came. The door was flung open, and Mary saw, first, a single red rose lying on the threshold. 'Ah, so *that's* what it's like to see red!' Mary thought to herself. Finally the curiosity that had burned within her breast for so long was satisfied. And so it went with all the other colours, as she explored the amazing world of colour outside of her black and white room. Whether or not knowledge of colours brought Mary lasting peace of mind is another story; but, for that day at least, Mary's curiosity was satisfied.

The above story expresses the following, seemingly innocuous thesis:

Phenomenal Curiosity (PC): Before she's experienced red, Mary could be curious about what it's like to see red, such that her curiosity cannot be (entirely) satisfied from reading physical science but can be satisfied when she experiences red for the first time.

In spite of its seeming innocence, I will suggest that PC can play a powerful dialectical role in debates over the knowledge argument (Jackson 1982). In what follows I will argue:

- that two prominent physicalist responses to the knowledge argument – the ability hypothesis and the phenomenal concept strategy – cannot account for PC,
- that a prominent anti-physicalist hypothesis, namely Revelation, can account for PC.

The upshot of this, or so I will argue, is to cast doubt on the middle ground between anti-physicalism and moderate forms of physicalism.

## 1. Phenomenal Curiosity and the Ability Hypothesis

Proponents of the ability hypothesis (Nemirov 1980, Lewis 1988) accept that Mary gains new knowledge when she sees red for the first time, but they deny that that knowledge is *propositional*; it is rather *know how*, consisting in the possession of certain abilities, e.g. the ability to imagine and remember red, and to distinguish things that are red from things that are not red. The problem for proponents of the ability hypothesis with respect to PC is that the mere gaining of abilities does not amount to a satisfaction of curiosity. Curiosity is

essentially a yearning for certain information; satisfaction of curiosity essentially involves the relief of gaining that information. The mere fact that I come to be able to do something that I couldn't previously do – wiggling my ears, say – does not entail that I have any new information, and hence does not entail that my curiosity is satisfied. The fact that Mary is able to do some new things when she sees red for the first time does not entail that she gains any new information, and hence does not entail that her curiosity is satisfied.

Of course, gaining new abilities *may* come with gaining new information. Gaining the ability to speak Chinese comes along with lots of information about what Chinese symbols represent. But the ability hypothesis does not specify which kinds of information might come along with gaining of the abilities that constitute 'knowing what it's like to see red', and hence accounting for what information might be gained would take us *beyond* the ability hypothesis as it has thus far been proposed.

Moreover, there is an obvious reason proponents of the ability hypothesis don't talk about Mary gaining new information: there is a worry that any such new information would be information about the non-physical nature of the experience, given that Mary already had all the information about the physical nature of the experience in her room. Indeed, this is the whole motivation for the ability hypothesis: avoiding the threat of what Lewis (1988) calls 'phenomenal information' by accounting for the change in Mary in terms of mere abilities.

One might object that it's obvious what kind of information is gained when one learns what it's like to see red: it's precisely the information that *this is what it's like to see red*. However, on the ability hypothesis, learning what it's like to see red just is a matter of gaining new abilities. Either the gaining of those abilities involves the gaining of new information or it doesn't. If it doesn't, then the ability hypothesis entails that learning what it's like to see red does not in fact involve gaining new information, and hence is inconsistent with PC; by definition, one's curiosity can be satisfied only if one gains new information. If Mary's new abilities do come with new information, on the other hand, then this undermines the whole point of ability hypothesis, which is to avoid the threat of phenomenal information by accounting for the change in Mary in terms of mere abilities.

## 2. Phenomenal Curiosity and the Phenomenal Concept Strategy

According to the phenomenal concept strategy (PCS), what Mary gains when she experiences red for the first time is *a new concept*; more specifically, she gains a phenomenal concept of red experiences (Loar 1990). Phenomenal concepts are those concepts we deploy when we think about experiences in terms of what it's like to have them. According to the phenomenal concept strategy:

- *Conceptual Dualism* – You can't move a priori from the fact that somebody is in some physical state to the fact that they are in some conscious state, or vice versa, simply because physical and phenomenal concepts are such different kinds of concept,
- *Metaphysical monism* – Nonetheless, every phenomenal concept co-refers with some physical concept.

Phenomenal concept strategists are typically happy to say that Mary gains new knowledge when she sees red for the first time. But that new knowledge does not involve learning about a new feature of reality; it wholly consists in gaining a different way of thinking about a property Mary already knew about. Mary already knew that people instantiate a certain pattern of neural firing when they see red things; post-release Mary can know that people have *this experience* when they see red things; but 'this experience' (i.e. red experience conceived of under a phenomenal concept) is just a new concept that refers to the pattern of neural firings she already knew about.

The problem for proponents of the phenomenal concept strategy with respect to PC is that the mere gaining of a new concept does not amount to a satisfaction of curiosity. Again, curiosity is essentially a yearning for certain information; satisfaction of curiosity essentially involves the relief of gaining that information. Merely gaining a new label for something I already knew about does not amount to the gaining of new information.

For some phenomenal concept strategists (Levin 2002), phenomenal concepts do provide some information about their referents. For example, Janet Levin defends an account of phenomenal concepts as providing information about resemblances between experiences, e.g., someone possessing a red phenomenal concept and an orange phenomenal concept can know a priori that red and orange experiences resemble. However, on all such forms of the phenomenal concept strategy thus far defended, the information provided by phenomenal concepts is structural information that Mary would already know about in her black and white room. This is not a coincidence. The whole point of PCS is to avoid the threat of non-physical properties, by holding that Mary just gains new concepts for properties she already knew about. If the PCS proponent concedes that phenomenal concepts provide information that Mary didn't know about in her black and white room, there is a worry that this new information would concern the non-physical nature of colour experience, given that Mary already had all the information about the physical nature of colour experience. In other words, it is for good reason that all extant PCS strategists hold either that phenomenal concepts reveal no information about their referents, or they reveal information which can also be gleaned from the physical facts that Mary would already have in her black and white room. In either case, the PCS strategist is unable to account for PC.

When I've raised this objection to PCS strategists, they have often replied: 'Well Mary's curiosity is satisfied because she learns *what it's like see red*. What's the problem?'. The problem is that, if the PCS strategy were true, Mary would *not* have had her curiosity satisfied. To have your curiosity satisfied, you have to gain new information, information you were previously yearning for. But if PCS is true, Mary does not gain any new information, she just gains a new way of referring to something she already knew about; and merely gaining a new way of referring to something you already knew about cannot, in itself, satisfy your curiosity.

Perhaps the proponent of PCS would object to the way I'm defining 'information.' There are more or less fine-grained ways of defining 'information', depending on whether one is thinking of information in terms of sets of metaphysically possible worlds or sets of

epistemically possible worlds. We could call what Mary gains, according to PCS, 'new information' if we liked. But this doesn't change the fact that all Mary gains, according to PCS, is a new way of referring to something she already knew about, and merely gaining a new way of referring to something you already knew about cannot, in itself, satisfy your curiosity. The kind of 'information', if we want to call it that, Mary gains according to PCS is not the kind of information that could satisfy curiosity.

The proponent of PCS might object that indexical knowledge can satisfy one's curiosity. If Clare Smith has suffered from amnesia, she can know all of the objective facts about Clare Smith and Sara Jones without knowing which of these persons she is. Her curiosity might be satisfied when she finds out, 'Ah, so I'm Clare!'. Couldn't something similar be going with Mary?

We might find an analogue to this in the case of Mariana, which is Martine Nida-Rümelin's (1996) twist on the knowledge argument. Mariana has exactly the same life as Mary, except that, before she is released, she is given colour chips, allowing her to perceive colours but without knowing, say, that *this colour* is the colour of bananas, *this colour* is the colour of tomatoes, etc. In this way, Mariana will gain phenomenal concepts before she leaves the room. At this stage, she might wonder, 'I wonder which of these is red?'. When she is finally liberated from her room, this curiosity could be satisfied, 'Ah, so *that colour* (picked out under a phenomenal concept) is red.'

However, the normal case of Mary is not analogous to the case of Clare described above. In the normal Mary, she gains a phenomenal concept of red at the same time as she sees red for the first time. And so the curiosity she had in her black and white ('I wonder what it's like to see red?') is not analogous to the curiosity that Clare has as to whether some entity known in one way (via a 1<sup>st</sup> person indexical) is the same as some entity known in a different way. If Mary's knowledge is indexical, then it's equivalent to gaining a new way of pointing at something you already knew about. Again, merely gaining a new way of referring to something one already knew about cannot satisfy one's curiosity.

Whatever view one adopts of the knowledge argument, there is another change in Mary when she sees red for the first time: she has a new experience, which, for the physicalist is (or is grounded in) a new brain state. But noting this doesn't help either. The mere fact that Mary has something new happening in her body does not entail that she is able to access any new information, and therefore cannot in itself account for her curiosity being satisfied.

### 3. Phenomenal Curiosity and Revelation

The thesis of Revelation (Stoljar 2009: 115) is defined in various ways, but it is roughly the claim that introspection and/or phenomenal concepts (can, at least partially) reveal the essential nature of their referents. I will define it as follows:

Revelation: For any conscious experience, one knows the essence of C if and only if one knows what it's like to have C.

In general, Revelation is defined with reference to introspection or phenomenal concepts. Note that my definition does not do this, and the benefit of this is that we avoid a commitment to introspection and/or phenomenal concepts perfectly revealing the essences of their referents. Many phenomenal concepts, e.g., the generic concept of 'pain', are very rough and ready, and so will miss out much of the essence of any particular experience. Some Revelation theses focus on what David Chalmers calls 'direct' phenomenal concepts, i.e., phenomenal concepts formed when we attend to a particular experience and allow the content of the concept to be wholly determined by the act of attending to that experience. However, Gerrit Neels (2021) has argued, on the basis of the non-transitivity of colour perception, that even here the phenomenal concept in question may not fully capture the essence of its referent.

Having said that, it's clear that introspection does reveal to us quite a lot about what it's like to have our experiences. When I attend to my pain, I know a great deal about what it feels like, even if reflective thought doesn't perfectly capture all aspects of the character of my pain. In conjunction with this obvious fact, Revelation entails that introspection reveals a great deal about the essences of my experiences; because, according to Revelation, to know the essence of a given phenomenal property just is to know what it's like to have it.

In contrast to the ability hypothesis and PCS, a commitment to Revelation easily accounts for PC (at least assuming, which all positions thus far discussed do, that physical science can't teach you what it's like to see red). Coming to know what it's like to see red, on Revelation, is not just a matter of gaining new abilities or gaining a new concept, but a matter of gaining new information about the essential nature of red experiences, new information that wasn't available to her in her black and white room. The proponent of Revelation can hold that it is the gaining of this information that satisfies Mary's curiosity.

As well as the full Revelation thesis outlined above, what would also do the job is the following:

Strong Partial Revelation: For any given experience E, to know what it's like to have E is to know *something* of the essence of E, something that goes beyond what can be discerned from physical science.

As noted in the last section, some forms of PCS hold that phenomenal concepts reveal information about phenomenal properties, but information that *can* in principle be discerned from physical science. We could call such views 'Weak Partial Revelation.' These views cannot account for PC, as Mary would already have the information provided by phenomenal concepts while in her black and white room. Strong Partial Revelation, however, can account for PC, as it entails that Mary would gain some new information when she learns what it's like to see red, information she may have been curious about while in her black and white room. In what follows, I will focus on Revelation rather than Strong Partial Revelation, but everything I say about the former holds true of the latter.

As discussed above, the satisfaction of curiosity is essentially a matter of yearning for certain information; the curiosity is satisfied when the information is accessed. What we need to account for PC is a story about what information Mary lacked in her black and white room

and gained when she was released. Revelation and Strong Partial Revelation each provide this: the information Mary lacked and then gained concerns the essential nature of colour phenomenal properties.

#### **4. Against Moderate Physicalism**

I have argued that Revelation can account for PC, whereas neither the ability hypothesis nor PCS are able to do this. Assuming for the moment that PC is true, this constitutes a case for adopting Revelation over these other responses to the knowledge argument. This is significant, as the ability hypothesis and PCS are the most popular ways of protecting conventional physicalist hypotheses from the knowledge argument, whereas Revelation fits with anti-physicalist hypotheses (or at least hypotheses which are not conventionally physicalist) like dualism or Russellian monism. Thus, assuming PC, we have a case for anti/not conventionally-physicalist hypotheses over conventional physicalist hypotheses.

To spell this out a bit more, let us define a conventional physicalist hypothesis as one that is compatible with physical science – the kinds of facts that Mary new in her black and white room – revealing the complete essential nature of conscious experience. If all Mary gains when she leaves the black and white room are new abilities or new concepts, it can be consistently held that Mary already had a complete understanding of the metaphysical nature of colour experience when in her black and white room.

However, if, as Revelation (or Strong Partial Revelation) supposes, Mary gains new information about the essential nature of colour experience upon liberation, then we can no longer hold that she already had complete information about the metaphysical nature of colour experience when still in her black and white room. Hence, a conventional physicalist hypothesis is ruled out. Still available are dualist hypotheses, according to which phenomenal properties have a non-physical essential nature, and Russellian monism, according to which physical science reveals the causal role of physical properties but not their essential nature.

Digging a bit deeper, it's not accidental that extent conventional physicalist responses to the knowledge argument can't account for PC. Having your curiosity satisfied consists in feeling relief that you gain some information you were yearning for. But conventional physicalist hypotheses can't concede that Mary gains new (non-indexical) information when she leaves the black and white room, as this would rule out that physical science could provide the complete essential nature of colour experiences. This is why the conventional physicalist has to offer something less than new (non-indexical) information to account for Mary's change, e.g., new abilities or concepts. We can account for PC only if we accept that phenomenal concepts reveal new information that goes beyond the information provided by physical science, and this is incompatible with conventional physicalism.

What about the option of rejecting PC? There are two possible ways of doing this:

Option 1: Deny that Mary could be curious about what it's like to see red, such that her curiosity is satisfied when she learns what it's like to see red.

Option 2: Hold that Mary could learn what it's like to see red (and thereby have her curiosity satisfied) while still in her black and white room.

I submit that the first option is deeply at odds with our everyday conception of our epistemological relation to our experiences. We do feel we can be curious about what it's like to have new experiences. Sometimes this curiosity can be satisfied by being told how the experience you haven't yet had resembles experiences you have had: David Lewis (1988) gave the example of being told that Vegemite resembles marmite. But if the new experience doesn't especially resemble any other prior experience, then you have to actually have the experience to have any your curiosity about it satisfied.

Perhaps one could hold that, although it's true as a general rule that you have to actually have an experience to know what it's like, sufficient neuroscientific knowledge could in principle also reveal to us what it's like to have any given experience, and thereby fully satisfy our curiosity regarding experiences we've never had. This would be to adopt Option 2, which would put one in the more radical camp of physicalists who deny even a conceptual distinction between the mental and the physical.<sup>1</sup>

Let us define 'moderate physicalism' as having the following two characteristics:

*Robust Realism* – Does not address the knowledge argument by holding that we are radically in error in our ordinary ways of conceiving of our experiences (i.e., rejects 'illusionist' theories of consciousness).

*Compromise* – Concedes that Mary learns something when she sees red for the first time (e.g., cashed out in terms of gaining new know how, or a new concept).

Rejecting PC means rejecting moderate physicalism so defined. It requires accepting one of the following three theses:

- our conviction that our curiosity can be satisfied when we learn what it's like to have a new experience is an illusion,<sup>2</sup>
- our conviction that there is something that it's like to see red is an illusion,
- a congenitally blind neuroscientist could learn what it's like to see red from reading neuroscience.

The first and second options give up Robust Realism; the third option gives up on Compromise.

---

<sup>1</sup> A middle way option would be to retain conceptual dualism, but to hold that we are only ever curious about structural information, or resemblances to experiences we already have, information which in principle Mary would have in her black and white room. However, such a view would have to deny the highly intuitive thesis that Mary could be curious about what it's like to see colour, which I take to be a cost. In support of this, Knut Nordby is a colour scientist who, due to having missing cones in his eyes, only experiences black and white and shades of grey, and who in his writing (Nordby 2007) expresses curiosity about what it's like to see colour. Indeed, his writing inspired my account of Mary's pre-release attempts to imagine colour.

<sup>2</sup> Or at least (see last footnote), our conviction that somebody (Mary or Nordby) could be curious about what it's like to see red, and have their curiosity satisfied when they find out.

## Conclusion

Reflecting on Mary's curiosity casts doubt on moderate forms of physicalism. Either we accept Phenomenal Curiosity, and are led to views which oppose conventional physicalism, such as dualism or Russellian monism, or we deny Phenomenal Curiosity, and are led to illusionism, or at least the denial of even a conceptual distinction between the mental and the physical. The middle ground of moderate physicalism, I suggest, is not sustainable.

Jackson, F. 1982, Epiphenomenal Qualia. *Philosophical Quarterly* 32: 127–136.

Levin, J. 2002. Is conceptual analysis needed for the reduction of qualitative states? *Philosophy and Phenomenological Research* 64: 571–591.

Lewis D. 1988. What Experience Teaches. *Proceedings of the Russellian Society* 13: 29–57.

Loar, B. 1990. Phenomenal States. *Philosophical Perspectives* 4: 81–108.

Neels, G. 2021. Goff's revelation thesis and the epistemology of colour discrimination. *Synthese* 199: 14371–82.

Nemirow, L. 1980. Review of T. Nagel, *Mortal Questions*. *Philosophical Review* 89: 473–77.

Nida-Rümelin, M. 1996. What Mary Couldn't Know. In *Conscious Experience*, ed. T. Metzinger, 219–41. Paderborn: Schöningh/Imprint Academic.

Nordby, K. 2007. What is this thing you call color: can a totally color-blind person know about color? In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, eds. T. Alter & S. Walter, 77–86. New York: Oxford University Press.

Stoljar, D. 2009. The Argument from Revelation. In *Conceptual Analysis and Philosophical Naturalism*, eds. D. Braddon-Mitchell and R. Nola, 112–38. Cambridge: MIT Press.