Philos Stud
DOI 10.1007/s11098-013-0110-1

Why it doesn't matter to metaphysics what Mary learns

Robert Cummins · Martin Roth · Ian Harmon

Abstract The *Knowledge Argument* of Frank Jackson has not persuaded physicalists, but their replies have not dispelled the intuition that someone raised in a black and white environment gains genuinely new knowledge when she sees colors for the first time. In what follows, we propose an explanation of this particular kind of knowledge gain that displays it as genuinely new, but orthogonal to both physicalism and phenomenology. We argue that Mary's case is an instance of a common phenomenon in which something new is learned as the result of exploiting representational resources that were not previously exploited, and that this results in gaining genuinely new information.

Keywords Knowledge argument · Physicalism · Phenomenology · Representation

1 The knowledge argument

> Mary, as everyone knows, was raised in a black and white environment. She learned all the physical science there is to know about color vision. But she

R. Cummins
Department of Philosophy, University of California at Davis, 1216 Dorchester Dr., Champaign, IL 61821, USA
e-mail: rowbear.cummins@gmail.com

M. Roth
Department of Philosophy and Religion, Drake University, 208 Medbury Hall, 2507 University Ave., Des Moines, IA 50311, USA
e-mail: martin.roth@drake.edu

I. Harmon (✉)
Department of Philosophy, University of Illinois at Urbana-Champaign, 105 Gregory Hall, 810 S. Wright St., Urbana, IL 61801, USA
e-mail: iharmon@illinois.edu

Published online: 27 February 2013

> didn't know what it is like to see red until she emerged from her restricted environment. When she emerged, she learned something she didn't know before, namely, what it is like to see red. She says (or thinks): *that is what it is like to see red.* Since she already knew all the relevant physical science, it follows that what she learned is not a physical fact. (After Jackson 1982 and 1986)

The Knowledge Argument has not persuaded physicalists, but neither have their replies dispelled the intuition that Mary gains genuinely new knowledge. This is as it should be, since, regardless of the plausibility of physicalism, it is prima facie suspicious to derive metaphysical conclusions from epistemological premises, premises about what can or cannot be known under various circumstances. The Knowledge Argument moves from what we can know only in a certain way to what there is. For those of us who look askance at such arguments, what makes the Knowledge Argument interesting (still) is precisely the challenge of accounting for how the knowledge Mary gains can be genuinely new and informative, and yet not constitute a threat to physicalism. There has been no shortage of physicalist attempts to answer this challenge, of course, but we think the arguments of dualists and physicalists alike have rested on flawed assumptions about the resources necessary to access physical information.[1] We argue that Mary's case is an instance of a common phenomenon in which something new is learned as the result of exploiting representational resources that were not previously exploited. Exploiting representational resources in this way allows for genuinely new information gains, but such gains have no metaphysical implications. Moreover, the possibility of such gains reveals that the issues at stake in the Knowledge Argument concern quite general features of representation and cognition, not features unique to first-person phenomenology.

2 Theoretical background

The account we will present of the knowledge Mary gains on emergence from her black and white environment depends on four assumptions:

A. *Representational theory of content* If Φ is the content of some mental state M of S, then M consists in part of a representation R whose content is Φ.
B. *Representational pluralism* There are multiple representational schemes, each with their proprietary content types and representational targets. These are

[1] For a survey of such replies, see Alter (2005). As we are going to use 'knowledge,' knowledge involves states with representational content. While we recognize that there may be accounts of Mary's knowledge gain that do not characterize the gain in terms of states with content—versions of the abilities reply (Lewis 1983, 1988; Nemirow 1990) may be examples—we are not going to discuss such views here. While we think there are compelling reasons to accept that Mary's genuinely new knowledge involves states with representational content, our aim isn't so much to defend this claim as to show why characterizing her genuinely new knowledge in terms of states with content (a) has no interesting metaphysical implications and (b) does not turn on features that are peculiar to first-person phenomenology.

generally not inter-translatable. Most of these are non-propositional in the sense that they are not candidates for truth-conditional semantics.[2]

C. *Psychological representational pluralism* The mind employs multiple representational schemes. These are generally not inter-translatable.
D. *Epistemological representational pluralism* A great deal of knowledge involves the exploitation of a diversity of representational schemes, both internal and external.

We take A to be relatively uncontroversial in the present context. While there are certainly those who reject the representational theory of content (Brooks 1991; van Gelder 1996), this is generally because they reject appeals to content generally as a strategy for understanding the dynamics of the mind/brain. Whatever the merits of that position, it is not likely to be attractive to friends of the Knowledge Argument, or to anyone who shares the intuition that Mary learns something new when (and because) she sees color for the first time.

Assumption B, though often ignored in practice, is quite generally acknowledged in principle. No one thinks that pictures, for example, are candidates for truth-conditional semantics, so depiction is clearly different than language, and the two are not inter-translatable. We can describe a picture, and we can construct a picture from a description, but this is not translation in any sense, however loose. The same point can be made about scale models, partitioned activation spaces, maps and sound recordings. A consequence of this is that one cannot express the content of such representations in language. This does not make these contents ineffable unless one simply means by this that they cannot be translated into language. They are unproblematically expressed and cognized in their own proprietary formats.[3] Pictures can be symbolically encoded, of course, as, e.g., pairs of gray-scale and position values. But a symbolic encoding of a picture of Beethoven (Fig. 1) does not depict Beethoven (or anything else), and its semantics does not overlap that of the picture. The encoding is about gray-scale and position pairs; the picture is about Beethoven's appearance. Because we have two very different schemes here, the *processing* is very different as well. Evolution has equipped us to process images, but it takes sophisticated software and a computer to process its pixel encoding. We can see immediately that the following are pictures of the same man:

Fig. 1 Two pictures of Beethoven

[2] See Haugeland (1991), Fodor (2007), Heck and Richard (2007), and Cummins (2010) for a sustained discussion and defense of theses along these lines.

[3] The words 'form' and 'format' are used here as a generalization of what is generally called syntax. Syntax is what we call the form or format of a structured propositional representation—e.g., a sentence.

Fig. 2 A man riding a bike

This is a non-trivial task for a system with only the pixel encoding. Indeed, even a resolution preserving resizing or a rotation of the same picture poses a substantial computational problem, since the encodings will be different.

Assumption C is, or ought to be, as uncontroversial as B. As Kant famously pointed out, percepts are not, and do not translate into, propositional thoughts. Seeing is not believing, or at least not believing a proposition. One often does, of course, believe that something is the case on the basis of what one sees. We even say that, e.g., we can see that the man on the bicycle is wearing a bowler hat (Fig. 2).

But this should be treated with care: the visual percept, like the picture, doesn't express a proposition. It is rather a representation—a depiction in this case—that allows us to infer a proposition: given that the visual percept (or the picture) is accurate, the man riding the bicycle is wearing a bowler hat. 'See', 'hear' and 'feel' can be misleading in this respect, because they can take propositional complements: Cantor allowed us to *see* that there are as many even natural numbers as there are natural numbers. Metal roofs make it all too easy to *hear* that it is raining. And one often *feels* that one's critics do not understand one's arguments. None of this should tempt us to think that percepts are expressions of propositions, or that they can be cognitively processed in the way expressions of propositions can be processed.

This, of course, is not a special property of percepts. In this respect, they are like the examples discussed under B. Percepts are like pictures, maps, graphs, recordings, and scale models: all are representational, but none of them are in the business of expressing propositions, and therefore are not candidates for truth-conditional semantics. They are, rather, more or less accurate representations of their targets. Moreover, they are all multi-dimensional, and accuracy along one dimension often comes at the cost of inaccuracy along another.[4]

Assumption D—epistemological representational pluralism—is the most controversial of the four assumptions, for, when put together with A through C, it yields the consequence that one can have knowledge that is represented in a picture, or model or percept that cannot be verbally expressed. One can know the street and intersection structure of London in virtue of having a cognitive map, or a paper or electronic map. One can know what Beethoven looked like in virtue of having a picture of Beethoven or a visual image generated on demand from memory.

[4] We also take it to be obvious that there is a good deal of diversity of format between perceptual modalities, and the science seems unequivocal that there is a good deal of diversity within a given modality as well.

Because of D, we are not happy with the way the terms 'knowledge' and 'fact' are deployed in the Knowledge Argument. Knowledge is generally taken to be a propositional attitude, or to involve a propositional attitude, the propositional content of which is what is known. Insofar as such contents can be captured in sentences, however, propositional contents are plausibly thought to be the contents of linguistic schemes of representation.[5] We think there is a good deal of knowledge that is not expressible in sentences and does not involve belief (when belief is construed as a propositional attitude), and it is plausible to suppose that such *non*-propositional knowledge plays a large role in science, everyday life, and the cognitive economy of the brain: graphs, pictures, images, scale models, diagrams, maps, and partitioned activation spaces and all manner of percepts are all chock full of more or less accurate information about the world, and none of these express propositions or are candidates for truth-conditional semantics.

We have correlative worries about 'fact'. Presenting or listing *the facts* generally means expressing a set of propositions. But no proposition has (as opposed to encodes) the content of a scale model or a partitioned activation space. Thus, on a familiar understanding of the facts, having all of the facts may not be the same as having all of the information or knowledge. On a standard philosophical construal of knowledge and facts—a propositional construal—it is, therefore, possible to know all the facts and still be missing an unbounded amount of solid scientific and common sense information about physical reality.

In what follows, we propose to use 'fact' in the usual way, as something corresponding to a true proposition. Since non-propositional representations come in various degrees of accuracy, often along many, possibly competing, dimensions, one's cognitive grip on reality will, likewise, be more or less accurate, or more or less effective, along many, possibly competing, dimensions. No flat map can accurately represent distances in a hilly terrain. It doesn't follow from this that a Roman cab driver who has memorized the map of Rome doesn't know how Rome is laid out, or how far it is from the central train station to the Coliseum. But it *does* follow that s/he does not know how far the Coliseum is from the central train station *simply in virtue of having the map.* Where it is important to accommodate the possibility of non-propositional knowledge, we will substitute 'cognitive gain' for 'knowledge gain', and 'has all the relevant information about T' for 'knows all the facts about T'.

3 Lingua and Mary

> Lingua, like Mary, was raised in a black and white environment. But Lingua's environment was also picture-free. She learned all the physical science there is to know about faces, and read descriptions of famous people, including

[5] According to both Fregean and Russellian accounts of propositions, propositions are significantly structured entities that have either senses (Fregean) or objects and properties (Russellian) as constituents. Each account reveals intimate parallels between propositional structure and sentential structure, and some philosophers (e.g., King 2007) hold that the structure of propositions is inherited directly from the structure of sentences. We think such parallels should come as no surprise.

> Beethoven. But she didn't know what Beethoven looked like until she emerged from her picture-free environment. When she was shown a black and white photo of Beethoven, she learned something she didn't know before, namely, what Beethoven looked like. She says (or thinks): *that is what Beethoven looked like.* Since she already knew all the relevant physical facts, it follows that what she learned is not a physical fact.

We take it this to be a transparently bad argument against physicalism. Understanding why it is bad, and why it is analogous to the Knowledge Argument, is the burden of the next two sections.

To begin, consider the following:

(a) Beethoven looked like this: 

Taken as a whole, (a) is an instance of what we will call a *content-demonstrating* speech act. Content-demonstrating speech acts have two basic components: a linguistic component containing a demonstrative (e.g., 'Beethoven looked like this') and an introduced representation [e.g., the black-and-white picture in (a)] the content of which is introduced by the demonstrative.[6] Other examples include:

- Pointing at a map and saying, "That is the layout of central Rome."
- Turning on a recording and saying, "This is Bach's famous chaconne for solo violin."
- Pointing at the Wikipedia animation in the entry on crankshafts and saying, "That is how you convert circular motion to linear motion."

The linguistic component of a content-demonstrating speech act can introduce a variety of different representations, including pictures, maps, models and graphs, as well as further linguistic representations (to name a few). As we will understand them, the demonstrative-containing linguistic component functions to specify a *target* for the representation introduced, and the result will be true just in case the representation is accurate of the specified target. In the case of (a), the truth condition can be rendered along the following lines:

[6] Stanley and Williamson (2001), while not using our terminology, posit content-demonstrating speech acts in their account of knowing how. On their view, knowing how to X entails knowing a proposition that contains a way of X-ing. They allow that a way of X-ing may be introduced into a proposition by way of demonstratives. In one of their examples, Hannah comes to know that some way is a way for her to ride a bicycle when Susan points to John and says, "That is a way for you (Hannah) to ride a bicycle." Contrary to our view, Stanley and Williamson do not think this commits them to positing non-propositional knowledge [cf. Stanley (2011)].

(b) The picture introduced is an accurate representation of what Beethoven looked like.[7]

We think it is obvious that someone could know (b) without having a clue what Beethoven looked like. Blindfolded people can point to the picture and say, with complete confidence, "Beethoven looked like this," provided that they have been reliably informed that the thing they are pointing to is a (reasonably accurate) picture of Beethoven. We also think it is obvious why knowing that this truth condition is satisfied is not enough to possess information concerning what Beethoven looked like: to possess information concerning what Beethoven looked like, you need access to the picture's content, and since the content of the picture introduced is not part of the specification of the truth condition, knowledge of the truth condition does not provide the relevant access.

The significance of this point may be obscured, however, by the thought that (b) can be exchanged for a specification that *does* include the content of the picture. For example, consider the following case:

(a1) JFK was killed like this: he was shot in the head.

Rendering the truth condition for (a1) along the lines of (a), we get

(b1) The sentence demonstrated is an accurate representation of how JFK was killed.

As with (b), it is obvious that one can know (b1) without knowing how JFK was killed. Since access to the content of the representation demonstrated ('he was shot in the head') is required to know how JFK was killed, and, since (b1) does not provide this access, knowing (b1) is not enough for knowing how JFK was killed. In this case, however, we can replace (b1) with something that would provide access to the relevant content, namely:

(b1*) A shot to the head killed JFK.

Clearly, (b1) obtains if and only if (b1*) obtains, so it may be tempting to conclude that the availability of (b1*) renders the lack of access to the demonstrated content provided by (b1) epistemologically insignificant. If we assume that there is nothing special about the present example in this regard, then we would have a *prima facie* reason to expect that the Beethoven case could be given a similar treatment.

We agree that the availability of (b1*) renders the lack of access provided by (b1) epistemologically insignificant, but the explanation of (b1*)'s capacity to do this also shows us why a similar treatment is *not* possible in the Beethoven case. Because the representation demonstrated in (a1) is a *linguistic* representation ('he was shot in the head'), there is a straightforward way to incorporate the content of this representation into a specification of a truth condition. By contrast, the content of the picture demonstrated in (a) *couldn't* be a part of a truth condition. Because sentences and pictures are semantically disjoint, there is no linguistic specification

[7] Accuracy can be graded, of course, and we will want to allow for (A) being true even if the picture introduced is not perfectly accurate. The point is tangential to the main argument, however, so we will ignore this complication here.

that has the content of the picture; *a fortiori*, there is no linguistic specification of a truth condition that could "load" the content of the picture into the truth condition. This is simply a consequence of representational pluralism (assumption B): pictures and sentences are not inter-translatable because there is no transformation from one to the other that is information/content preserving. To repeat: we can describe a picture, and we can draw from a description, but these are not translations; information is lost in converting one format to the other. The police artist who incrementally modifies a picture in response to input from an eyewitness is creating a different kind of representation with a different kind of content in response to error signals from the witness.[8]

With this in mind, let's return to Mary's case. Consider the following:

(a2) Red objects look like this[9]: 

Rendering the truth conditions for (a2) along the lines of (a) and (a1), we get:

(b2) The image introduced is an accurate representation of what red objects look like.

As was the case with (b), it should be clear that a person could know (b2) without having a clue what red objects look like. Mary, sequestered in her black and white environment, can be talking on the phone to her friend Paul, who tells her that he is holding an image of a red rectangle. Mary can then say of the image, "That is what red objects look like." However, because the representation introduced in (a2) is not a linguistic representation, there is no sentence or set of sentences that will allow Mary to access the content of the image; to access the content, she actually needs the image (or something else that is red). Furthermore, it seems clear that having access to the image would enable Mary to gain genuinely new information about what red things look like.

(a) and (a2) make cognitive gains possible, but this possibility requires being in a position to exploit a certain kind of representational resource, a resource that is only available for exploitation under special circumstances. In the case of (a), for example, you have to have the picture, it cannot be dark, it cannot be too small or too big given your distance from it, etc. In the case of (a2), you have to have color vision and an available red object at a reasonable distance in reasonable light, etc. Mary, the story assumes, has the capacity to exploit the representational resource in question: she has normal color vision, but she just does not have the opportunity to exploit it. The same applies to someone who has pictures of Beethoven at their disposal, but is in the dark. When the lights go on, they can learn something, viz., what Beethoven looked like. In both cases, the cognitive gain is explained by a

[8] This is a kind of search by gradient descent in error space analogous to having someone find a hidden object in response to "warmer" or "colder" signals from someone who knows the location.

[9] Due to production limitaions, the square is not colored in the print version of the paper. The square is colored red in the online version of the paper, and should be understood as being red by the reader. Mary, of course, knows it is the color—not the shape—that is at issue.

change in circumstance that makes it possible to exploit a representational resource that could not be exploited previously.[10]

4 Metaphysics and genuinely new cognitive gain

The key to understanding how the respective gains of Lingua and Mary can be *genuinely new* is noticing that what Lingua and Mary learn are not *facts,* if we assume that a fact is something expressible by representations that are true or false. Knowing all the physical *facts*, in this sense (assuming counter-factually that this is possible), would leave Lingua as ignorant of Beethoven's appearance as it leaves Mary ignorant of what red things look like. What both require is information of a kind that their limited representational opportunities renders inaccessible. Mary is limited by lack of a target for her capacity for color vision; Lingua is limited by the absence of pictorial representations generally. The Knowledge Argument assumes that there is a way to limit representational resources, or to limit opportunities to exploit them, without limiting epistemological access to the physical. In Lingua's case, the lack of pictorial representations prevents her from accessing information about Beethoven's appearance. This, perhaps, does not limit her access to the *facts* (physical or otherwise), but the example shows that it certainly does limit her access to information about matters whose status as physical seems uncontroversial.

Defenders of the Knowledge Argument are not alone in assuming that limiting non-propositional representational resources (or opportunities to exploit them) does not limit epistemological access to the physical. Indeed, we think that the "new knowledge/old fact" response to the Knowledge Argument (Horgan 1984; Tye 1986; Loar 1990; Lycan 2003) is committed to the same assumption. According to that strategy, upon release Mary acquires or exercises concepts that she could not have acquired or exercised pre-release. These concepts allow her to represent "old facts" under what amounts to new modes of presentation of those facts. The gain accruing to Mary is thus claimed to be like the gain accruing to Oedipus: Oedipus *discovers* that he is married to his mother, even though (a) Oedipus already knows that he is married to Jocasta, and (b) the fact that he is married is his mother is identical to the fact that he is married to Jocasta. The knowledge gain is possible because although the woman was his marriage-object under one description or mode of presentation, she was not his marriage-object under the other (Lycan 2003). Applying this kind of analysis to Lingua's case, one would hold that what Lingua acquires is, as it were, a new take on a fact already in her possession, an idea that is evidently motivated by a desire to preserve the crucial premise of the Knowledge Argument that Lingua knew all the relevant physical facts before emerging from her picture free environment.

Some critics of the new knowledge/old fact strategy argue that the robust nature of Mary's gain makes it implausible to think that what Mary learns is something she

[10] Lingua can get a good verbal description, but there would still be a substantial cognitive gain when the picture becomes available. This is why you want the police to have a picture of the suspect rather than a verbal description.

already knew under another guise.[11] We agree with these critics, and we think the point applies equally to Lingua. But Lingua's case makes it clear, in a way that Mary's case perhaps does not, that we shouldn't be tempted to wed the issue of physicalism to whether the new knowledge/old fact reply can accommodate robust knowledge gains. As we noted in Sect. 2, on the standard reading of 'knowing the facts', knowing the facts amounts to knowing a set of propositions. The new knowledge/old fact reply is thus committed to saying that what Mary gains is a form of propositional knowledge. However, once we abandon the assumption that limiting non-propositional representational resources (or opportunities to exploit them) does not limit epistemological access to the physical, it should be obvious that salvaging physicalism does not require that Lingua's gain involve propositions. This simply follows from representational pluralism. And since exploiting a previously unexploited scheme of representation introduces novel contents, our view provides a straightforward explanation of how both Mary's and Lingua's information can be genuinely new and yet pose no threat to physicalism.

The concern that the new knowledge/old fact strategy does not account for the seemingly robust nature of Mary's new knowledge has led Tye (2009) to abandon this strategy in favor of the idea that Mary's cognitive gain is partly constituted by her becoming *acquainted* with red. Prior to her release, Mary knows all the physical facts *about* red; this is conceptual knowledge, or knowledge-that. However, according to Tye, there is a logically distinct kind of knowledge—acquaintance knowledge—that Mary lacks (p. 132). Acquaintance involves representations that are non-conceptual (p. 100), by which Tye means that the representations do not have concepts as constituents, and their contents are course-grained (p. 104).

While Tye's new view has certain affinities with the account of cognitive gain being offered in this paper, our account is importantly different from his. According to Tye, course-grained contents are Russellian singular propositions (p. 82). Moreover, while pre-release Mary is not acquainted with such propositions, pre-release Mary can entertain them (p. 98). However, because Russellian propositions are reasonably thought of as the proprietary targets of linguistic schemes of representation, Tye's account does not appear to explain Mary's cognitive gain in terms of accessing genuinely *novel* contents.

[11] Michael Tye puts the point this way:

> *What* Mary thinks is not new when she leaves her room. What is new is the *way* she is thinking what she is thinking. That isn't enough. What Mary knows before time *t* (the time of her release) is exactly the same as what she knows after time *t*. But if what she knows before and after her release is the same, she does not make a discovery in a really robust sense. This is counter-intuitive. Surely if anyone ever made a significant discovery, Mary does here. The proposal, in the end, is not convincing (2009, p. 55)

Notice, by the way, that if you (a) want to be a physicalist, (b) assume that limiting non-propositional representational resources (or opportunities to exploit them) does not limit epistemological access to the physical, (c) accept that Mary gains genuinely new knowledge after her release, but (d) deny that the new knowledge/old fact reply explains how Mary's knowledge can be genuinely new, then the abilities reply will look rather compelling. Because we hold that Mary's gain should be characterized in cognitive-cum-representational terms, with think the new knowledge/old fact reply is superior to the abilities reply. The important point here, however, is that the replies are equally flawed insofar as each assumes (b).

Historically, our access to the physical has depended essentially on the development of new representational resources (tomography, fractal geometry) and new opportunities to exploit them (fMRI, the Hubble telescope). Cognitive gains resulting from augmented representational resources or opportunities to exploit those resources are thus as commonplace as they are important, but such gains evidently have no deep metaphysical significance in themselves. While Mary benefits from an opportunity to exploit a previously unexploited resource when she emerges from her chromatic quarantine, the new information she gains has, *by itself*, no interesting metaphysical implications.[12]

5 What it's like

Conspicuously absent from our discussion of the Knowledge Argument so far is any mention of *what it's like*. There are two reasons for this: (i) our response to the Knowledge Argument rests on quite general considerations of representation and cognition, not on features peculiar to *what it's like*; (ii) while Lingua's case involves genuinely new information, the robust gain involved doesn't obviously have anything to do with phenomenology or *what it's like*; rather, the robust gain seems to have everything to do with a property of *Beethoven*. In this section we explore the implications (i) and (ii) have for the role *what it's like* plays in the Knowledge Argument. We'll start with (ii).

Once we appreciate that Lingua's robust gain has little to do with phenomenology, or *what it's like*, we should begin to take seriously that Mary's robust gain also has little to do with *what it's like*. Indeed, the significant gain seems to have everything to do with properties represented in vision, not *what it's like* to visually represent those properties. Note that in claiming this, we needn't deny that their respective gains come with additional information about *what it's like*. The issue here concerns the source of the *robust* gain, for, as far as responding to Knowledge Argument goes, even if information about *what it's like* to be in a perceptual state is distinct from the representational content of the perceptual state, information gained about *what it's like* is clearly relevant to the Knowledge Argument only if it can be reasonably supposed that information concerning *what it's like* is itself a source of robust gain.[13] In Lingua's case, it is plausible to suppose that the *robust* gain is exhausted by the information acquired about a property of Beethoven—what *he*

[12] We don't wish to legislate concerning what counts as metaphysics. Our point is simply that the fact of Mary's cognitive gain provides, by itself, no better reason to accept dualism than is provided by Lingua's cognitive gain.

[13] If information about *what it's like* to see red depends on information about what red things look like, as it surely does, then won't it trivially follow that the information gained about *what it's like* to see red is robust if the information about what red things look like is robust? Yes, but we can avoid trivializing the issue by reframing it as follows: is the information about *what it's like* to see red robust in a way that is not entirely parasitic on the robust gain involved in acquiring information about what red things look like? If the answer is no, then the information gained about *what it's like* to see red is robust in a merely derivative sense. We think this would be enough to show that information about what red things look like is doing the important work in this debate, not *what it's like* information. Of course, if *what it's like* aspects associated with perceptual states *just are* the representational contents of such states (e.g., Dretske

looked like. We think this is enough to place the burden of proof on those who would claim Mary's case is importantly different. The response cannot be simply that Mary's case involves new colors, whereas Lingua's does not. That would evidently beg the question, for the claim being made here is that the source of the *robust* gain—the gain that gets the Knowledge Argument off the ground in the first place—is exhausted by the content of the visual state, and the difference between this kind of content and the kind of content gained by Lingua is orthogonal to *what it's like*. Thus, short of being given some good reason for thinking that Mary's case is importantly different from Lingua's, an account of the robust gain involved in acquiring information about what red things look like is the only account we can be reasonably asked to provide.[14]

With that said, robust information gains concerning *what it's like* wouldn't be probative in any case. This is the important lesson to be learned from (i). Suppose we grant that exploiting a non-linguistic scheme to represent colors would not by itself account for Mary's information concerning *what it's like* to represent red visually, and suppose we grant that her new information concerning *what it's like* constitutes a robust gain. It should now be clear why the existence of such a robust gain would tell us nothing about the status of physicalism. To defeat the Knowledge Argument, we don't need to provide a detailed account of robust information gains involving *what it's like*. Instead, we need to show that genuinely new cognitive gains do not, *by themselves,* have interesting metaphysical implications. When Lingua emerges from her picture-free (yet still black and white) environment, Lingua exploits a previously unexploited scheme of representation and thereby comes to possess genuinely new information that she could not have possessed otherwise. This gain has no interesting metaphysical implications, and no one not already under the spell of the Knowledge Argument would suppose it has anything to do with significantly new *what it's like* information. Even if Mary's case does involve significantly new *what it's like* information, the lesson to draw is not that Lingua's case leaves out something important, something that is involved in Mary's case and is relevant to the issue of physicalism. The lesson to draw, rather, is that the issue relevant to the Knowledge Argument concerns genuinely new cognitive gains, and the issue of genuinely new cognitive gains is orthogonal to the issue of *what it's like*. Mary's genuinely new information about *what it's like* would imply no more about physicalism than is implied by Lingua's genuinely new information about

Footnote 13 continued
1995; Tye 1995), it is necessary that gains concerning the former are robust if gains concerning the latter are robust. While we have some sympathy for this position, we don't assume it here.

[14] Though we don't think the case has been made for showing that *what it's like* information gains are especially relevant to the Knowledge Argument, we think an "inner sense" account (e.g., Armstrong 1968; Lycan 1987), suitably adapted to accommodate the points we make about representation and knowledge, could provide a plausible model of such gains. Just as you can take a photo of a photo, so you can have a perception of a perception—this is what inner-sense is. The important points are this (1): on the inner sense account, we just have another case of an opportunity to apply a representational resource to a target not previously available, namely seeing red; (2) that resource is not (purely) propositional any more than vision is purely propositional. If it were purely propositional, Mary could just be told what it is like to see red. See Churchland and Churchland (1998) for an example of what an inner sense view developed along these lines might look like.

Beethoven. *What it's like* is a red herring when it comes to understanding the flaw in the Knowledge Argument, and, more generally, the flaw in deriving a metaphysical conclusion from premises about representational resources and opportunities to exploit them.

To get the anti-physicalist conclusion, the Knowledge Argument requires a premise specifying what resources are sufficient to represent the physical—an assumption that the history of science suggests is not going to be forthcoming from armchair thought experiments. The Knowledge Argument assumes that you can limit experience in certain ways without limiting information about the physical. This might seem plausible if you assume (i) that limiting experience isn't limiting representational resources or opportunities to exploit them, or (ii) that the physical must be fully representable with limited representational resources and opportunities. Assumption (i) is implausible and unsupported. Moreover, (i) is not available to the advocate of the Knowledge Argument, since that argument assumes that Mary gains new information on emerging from her chromatic quarantine, and this is surely because she is then able to exploit her ability to generate representations she could not previously generate. Assumption (ii), as we have seen, is hopeless in its full generality. It only looks plausible when it is limited to the claim that the physical is fully representable without recourse to something inherently first personal (as *what it's like* knowledge is assumed to be). But imposing that limitation would simply beg the question at issue. What it takes for *us* to know about the physical is as much a question about *us* as it is a question about the physical. No sensible physicalism is a thesis about what it takes to *represent* the physical. Nagel (1974) wisely refused to draw a metaphysical conclusion from the fact that some knowledge is inherently first personal. Instead, Nagel settled for an epistemological conclusion, viz., that there could not be a science of what is known only from the first person point of view. We are agnostic on the issue of the limits of science, but we applaud Nagel's unwillingness to move from a premise about what it takes for us to know something to a conclusion about the metaphysical status of the thing known. That way lies anti-realism, a territory where the issue of physicalism is moot at best.

If we are right about Mary's cognitive gain, then it is in good and plentiful company, no more and no less significant or special than Lingua's. Knowledge of what it's like may be epistemologically unique in some deep and important sense, and we can concede the possibility that some information might be available to us, now, given our cognitive and neural design and our current technology, only via the first person. But it won't follow from this that the information in question isn't information about the physical; it will follow only that we can get different kinds of information in certain ways and not others, and that is a conclusion we can derive from representational pluralism alone.

6 Conclusion

A satisfying response to the Knowledge Argument requires an account of the genuinely new cognitive gain Mary achieves when she emerges from her black and white environment. Our strategy is to assimilate this case of cognitive gain to a more

general and familiar kind of case, the case of exploiting a representational resource that was previously not exploited. You cannot take a photograph of a sunset if you do not have a camera. But neither can you take a photograph of a sunset if you do not have access to a sunset (or a suitable substitute target). You cannot trace a Sheppard drawing of Winnie the Pooh if you do not have tracing paper. But neither can you trace a Sheppard drawing of Winnie the Pooh if you do not have access to the drawing. You cannot taste a pineapple if your taste buds are numb. But neither can you taste a pineapple if you do not have a pineapple (or a suitable substitute stimulus). You cannot see red if you are blind, or in the dark. But neither can you see red if you do not have access to red things (or a suitable substitute). When we see Mary's cognitive gain as the result of exploiting a representational capacity that previously lacked an appropriate target, we see her cognitive gain as an example of a phenomenon that has interesting and little noticed epistemological implications, but no metaphysical implications at all.

This is as it should be. You cannot legitimately draw metaphysical conclusions from cases like these without a premise to the effect that everything with a given metaphysical status must be representable without the missing resources or opportunities to exploit them. But a premise of this kind—e.g., that all information about what is physical must be linguistically representable—is difficult to justify without begging the question. For it will always be open to the physicalist to say that cases like Lingua's, and the history of science, show that it takes more than language to achieve all the possible cognitive gains about things physical. We don't know what it will take. We are pretty sure no one else knows either.

References

Alter, T. (2005). The knowledge argument against physicalism. *Internet encyclopedia of philosophy*. http://www.iep.utm.edu/know-arg/.

Armstrong, D. (1968). *A materialist theory of the mind*. London: Routledge.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence, 47*(1–3), 139–559.

Churchland, P., & Churchland, P. (1998). *On the contrary*. Cambridge, MA: MIT Press.

Cummins, R. (2010). Representational specialization: The synthetic a priori revisited. In *The World in the Head* (pp. 194–209). Oxford: Oxford University Press.

Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: The MIT Press.

Fodor, J. (2007). The revenge of the given. In B. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 105–116). Malden, MA: Blackwell.

Haugeland, J. (1991). Representational genera. In W. Ramsey, S. Stich, & D. Rumelhart (Eds.), *Philosophy and connectionist theory* (pp. 61–89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Heck, R. (2007). Are there different kinds of content?" In B. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 117–138). Malden, MA: Blackwell.

Horgan, T. (1984). Jackson on physical information and qualia. *Philosophical Quarterly, 34*, 147–152.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly, 32*, 127–136.

Jackson, F. (1986). What Mary didn't know. *Journal of Philosophy, 83*, 291–295.

King, J. (2007). *The nature and structure of content*. Oxford: Oxford University Press.

Lewis, D. (1983). Postscript to "mad pain and martian pain." In *Philosophical papers* (Vol. 1). Oxford: Oxford University Press.

Lewis, D. (1988). What experience teaches. *Proceedings of the Russellian Society, 13*, 29–57.

Loar, B. (1990). Phenomenal states. *Philosophical Perspectives, 4*, 81–108.

Lycan, W. G. (1987). *Consciousness*. Cambridge, MA: MIT Press.
Lycan, W. G. (2003). Perspectival representation and the knowledge argument. In Q. Smith & A. Jokic (Eds.), *Consciousness: New philosophical perspectives* (pp. 384–395). Oxford: Oxford University Press.
Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 84*(4), 435–450.
Nemirow, L. (1990). Physicalism and the cognitive role of acquaintance. In W. G. Lycan (Ed.), *Mind and cognition* (pp. 490–499). Cambridge, MA: Blackwell.
Stanley, J. (2011). Knowing (how). *Nous, 45*(2), 207–238.
Stanley, J., & Williamson, T. (2001). Knowing how. *The Journal of Philosophy, 98*(8), 411–444.
Tye, M. (1986). The subjective qualities of experience. *Mind, 95*, 1–17.
Tye, M. (1995). *Ten problems of consciousness*. Cambridge, MA: The MIT Press.
Tye, M. (2009). *Consciousness revisisted: Materialism without phenomenal concepts*. Cambridge, MA: MIT Press.
Van Gelder, T. (1996). Dynamics and cognition. In J. Haugeland (Ed.), *Mind design II*. Cambridge, MA: MIT Press.