

Consciousness and Introspective Inaccuracy¹

Derk Pereboom, University of Vermont

Tentatively slated for *Appearance, Reality, and the Good: Themes from the Philosophy of Robert*

M. Adams, L. M. Jorgensen and Samuel Newlands, eds.

A Kantian perspective on the nature of introspective awareness, I contend, inspires a defense of a physicalist understanding of phenomenal states in the face of the most prominent arguments against it. Immanuel Kant claims that introspective representations -- those of *inner sense* -- are entities caused by the states they represent and are distinct from them, and they mediate the representational relationship between the subject and the introspected psychological states. As a result, the subject may not represent these states as they are in themselves.² I will argue that Kant's position yields a significant challenge to Frank Jackson's knowledge

¹ This paper was presented at *Metaphysics, History, Ethics*, a conference at Yale University in April 2005, in honor of Robert Adams, my dissertation advisor at UCLA. Thanks to Keith DeRose for his very fine comments at the session, and to the audience, especially Robert Adams, for high-quality questions and reflections. It was also presented in colloquia at the University of Auckland, the Australian National University, and the University of Alabama, and I'm grateful to audiences there for enlightening discussions. Thanks in addition to Kati Balog, Fiona Macpherson, Laurie Paul, Denis Robinson, David Kaplan, Adam Wager, Brian Weatherston, Sin yee Chan, and Hilary Kornblith for helpful comments and conversation. Special thanks are owed to Torin Alter, David Barnett, David Chalmers, David Christensen, Louis deRosset, Tyler Doggett, Mark Moyer, and Daniel Stoljar for extensive and valuable commentary, discussion, and correspondence. Research on this article was facilitated by a generous Visiting Fellowship in the Centre for Consciousness of the Research School of Social Sciences at the Australian National University.

² Immanuel Kant, *Critique of Pure Reason*, tr. Paul Guyer and Allen Wood, (Cambridge: Cambridge University Press, 1987), B152-4. Robert Adams suggests that Leibniz's views on perception provide at least as good a model (fill out note).

argument,³ and that it provides a response to those, like Joseph Levine and Robert Adams, who maintain that there is an explanatory gap between the physical and the phenomenal that the physicalist will have difficulty closing.⁴

1. Qualitative inaccuracy.

In Jackson's story, Mary has lived her entire life in a room that displays only various shades of black, white, and grey.⁵ She acquires information about the world outside, and also about the physical nature of the human being, by means of a black and white television monitor. By watching television programs Mary eventually comes to have knowledge of all of the physical information there is about the nature of the human being. (This complete physical knowledge might be conceived as complete microphysical knowledge, or as complete knowledge of any entity that is uncontroversially physical, or else as exhaustive factual knowledge of every entity that is wholly physically constituted -- each of these proposals might have its advantages and disadvantages.) But even if she knows all of this, Jackson contends, there is much she will not know about human experience. She will not know, for example, what it is like to visually

³ I contend that David Chalmers's zombie argument is also vulnerable to this strategy, although I defer development of this claim to another occasion (but see note 26).

⁴ Joseph Levine, "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly* 64, pp. 354 -61, and *Purple Haze* (Oxford: Oxford University Press, 2001); Robert Adams, "Flavors, Colors and God," in Adams, *The Virtue of Faith* (Oxford: Oxford University Press, 1987), pp. 243-62.

⁵ Frank Jackson, in "Epiphenomenal Qualia," *Philosophical Quarterly* 32 (1980), pp. 127-36, and in "What Mary Didn't Know," *The Journal of Philosophy* 83 (1986), pp. 291-95; cf. Thomas Nagel, "What is it Like to Be a Bat?," *The Philosophical Review* 83 (1974), pp. 435-50.

experience a ripe red tomato, and in particular, she lacks knowledge of what it is like to see red.

When she leaves the room and sees a red tomato, she comes to know for the first time – she

learns -- what it is like to see red. She gains knowledge, for the first time, of a particular

phenomenal property, or of a mental state that has this property – a *phenomenal state*.⁶ Thus

there are facts about phenomenal states that are not physical facts, and thus phenomenal states are

not completely physical. The core intuition underlying the knowledge argument is that if

someone who possesses complete physical knowledge does not *thereby* know some fact about a

phenomenal state, then that fact cannot be physical, and moreover, the phenomenal state cannot

be completely physical.⁷

Now consider the “old fact/new guise” response to the knowledge argument -- which I do not endorse, but the reply I develop can be understood as a successor to it.⁸ According to this

⁶ David Chalmers characterizes phenomenal properties as those that “type mental states by what it is like to have them,” “The Content and Epistemology of Phenomenal Belief,” in Q. Smith and A. Jovic (eds), *Consciousness: New Philosophical Perspectives* (Oxford, 2003). The “what it is like to have them” locution should perhaps be taken as a means of signaling to an audience what to look for as instances of phenomenal properties, which can then serve as paradigms, and not so much as a thorough descriptive characterization of this type of property.

⁷ One reply to the argument is that the reason that pre-emergence Mary lacks knowledge of phenomenal states is just that she is missing phenomenal concepts. In response, Daniel Stoljar strengthens the argument by specifying that pre-emergence Mary possesses all the phenomenal concepts, while she nevertheless lacks knowledge of how correct applications of phenomenal concepts are correlated with physical states; “Physicalism and Phenomenal Concepts,” forthcoming, *Mind and Language*; Stoljar tells a plausible story as to how Mary might come to fit this description (after acquiring the phenomenal concepts Mary suffers selective amnesia); David Chalmers makes a similar point in “The Two-Dimensional Argument Against Materialism,” forthcoming in *The Character of Consciousness* (Oxford: Oxford University Press, 2006). The resulting argument has a somewhat different focus. What I say in section 5 in reply to Adams is also a response to this argument.

⁸ Proponents of the “old fact/new guise” response include Terence Horgan, “Jackson on Physical Information and Qualia,” *Philosophical Quarterly* 32 (1984), pp. 147-52; Paul M. Churchland,

kind of response, Mary, when she is still in the room, does indeed know every fact about phenomenal states by virtue of her exhaustive physical knowledge, while what she is missing are only ways of introspectively representing those states, or as I will put it, *introspective modes of presentation* of those states.⁹ When she leaves the room and sees the red tomato, she comes to represent a phenomenal state, about which she already knew everything, by an introspective mode of presentation, with which she had never represented that phenomenal state while she was in the room. In this way, the *appearance* of Mary's coming to know a new fact can be explained without granting that she acquires new knowledge.

This sort of reply might be illustrated by various analogies. According to William Lycan, the difference between the introspective and the physical representations is akin to the difference between my use of 'I' and your use of 'you' to represent me in the representation of some fact about me.¹⁰ For example, consider:

(1) 'I weigh 190 pounds' (asserted by me)

(2) 'You weigh 190 pounds' (asserted by you)

"Reduction, Qualia and the Direct Introspection of Brain States," *Journal of Philosophy* 82 (1985), pp. 8-28; Robert Van Gulick, "Physicalism and the Subjectivity of the Mental," *Philosophical Topics* 13 (1985), pp. 51-70; Michael Tye, "The Subjective Qualities of Experience," *Mind* 95 (1986), pp. 1-17; Brian Loar, "Phenomenal States," *Philosophical Perspectives 4: Action Theory and Philosophy of Mind*, ed. James Tomberlin, (Atascadero, CA: Ridgeview Publishing Company, 1990), pp. 81-108; William G. Lycan, "What is the "Subjectivity" of the Mental?" *Philosophical Perspectives* 4, pp. 109-30.

⁹ I use the Fregean term 'mode of presentation' as a convenient nominalization, without intending the full Fregean theory. The claims made in this paper can generally be made in more neutral terms, or in terms of other theories of cognition and language.

¹⁰ William G. Lycan, "What is the "Subjectivity" of the Mental?"

You cannot represent the fact that I weigh 190 pounds by 'I weigh 190 pounds,' whereas I can represent this fact by means of that sentence. But suppose that you have knowledge of this fact, and represent it by 'You weigh 190 pounds.' Then there is no fact of which I have knowledge but you don't; the only fact to be known here is that DP weighs 190 pounds, and we both know it.

Although some find analogies of this sort sufficient to dislodge the knowledge argument, its proponents remain unconvinced.¹¹ To advance the debate, we need to explore why the argument has this residual force. Are there features of Mary's epistemic situation disanalogous with Lycan's example that might explain this force? Phenomenal states have characteristic phenomenal properties, and it is intuitive, for some, at least, that:

(i) both the physical and introspective modes of presentation represent these phenomenal properties as having a qualitative nature, and the qualitative nature that the introspective mode of presentation represents a phenomenal property as having is not included in what the physical mode of presentation represents it as having.¹²

It is also intuitive – again, for some -- that:

(ii) An introspective mode of presentation *accurately represents* the qualitative nature of

¹¹ A sophisticated reply along these lines is provided by John Perry, *Knowledge, Possibility, and Consciousness* (Cambridge Mass.: MIT Press, 2001); see also John Hawthorne, "Advice for Physicalists," *Philosophical Studies* 109 (2002), pp. 53-74; for someone who is not convinced by these sorts of accounts, see David Chalmers, "Imagination, Indexicality and Intensions," *Philosophy and Phenomenological Research* 68 (2004), pp. 182-90.

¹² Joseph Levine, in *Purple Haze*, accounts for the existence of the explanatory gap partly by the fact that "modes of presentation whereby we come into cognitive content with qualia are substantive and determinate" (p. 8), and that "there is real content to our idea of a quale," (p. 84). In what I am saying here I aim to explicate these kinds of intuitions; cf. Alex Byrne, "Review of *Purple Haze*," *The Philosophical Review* 111 (2002), pp. 594-7.

a phenomenal property. That is, an introspective mode of presentation represents a phenomenal property as having a particular qualitative nature, and the attribution of this nature to the phenomenal property is correct.

There is no uncontroversial way to characterize the qualitative nature that introspective modes of presentation present phenomenal properties as having. One option, inspired by John Locke, is to characterize this nature by way of *resemblance* to modes of presentation. Thus, in our example, we might say that Mary's introspective representation of her phenomenal-red sensation presents that sensation in a *what-it-is-like-to-sense-red* way, and it is intuitive that a qualitative nature that resembles this what-it-is-like mode of presentation is correctly attributed to the phenomenal property.¹³ Or, in deference to concerns about the cogency of such resemblance characterizations, one might say simply that *the phenomenal property is as it is introspectively represented*.¹⁴

¹³ Accepting a resemblance claim of this sort does not amount to endorsing a discredited *resemblance theory of representation*, as is sometimes suggested. For in accepting that the phenomenal property accurately attributed to the state resembles the introspective mode of presentation, one is not also accepting a resemblance account of how it is that the mode of presentation represents the phenomenal property. By analogy, one does not have to accept a resemblance account of photographic representation in order to accept the claim that photographs can resemble what they represent.

¹⁴ Note that the accuracy claim (ii) is weaker than what David Lewis calls *revelation*, which is: "...when I have an experience with quale Q, the knowledge I thereby gain reveals the essence of Q: a property of Q such that, necessarily, Q has it and nothing else does"; ("Should a Materialist Believe in Qualia," *The Australian Journal of Philosophy* 73 (1995), pp. 140-4; reprinted in Lewis's *Papers in Metaphysics and Epistemology*, (Cambridge: Cambridge University Press, 1999), pp. 325-31, at p. 328.) The accuracy claim is not that the essence of the quale is revealed in an (introspective) experience of the quale, but rather that the quale really has the qualitative nature this experience represents it as having. This is consistent with this experience not representing the complete essence of the quale.

Given these claims about what is intuitive, an advocate of the knowledge argument can account for its residual force in the following way. When Mary leaves the room and sees the tomato, she comes to believe that

(T) phenomenal redness has qualitative nature Q.

Qualitative nature Q is accurately represented introspectively by way of the *what-it-is-like-to-sense-red* introspective mode of presentation. But on the physicalist hypothesis, every truth about the qualitative nature that an introspective mode of presentation accurately represents a phenomenal property as having would need to be derivable from what the relevant physical mode of presentation represents this property as having. However, (T) is not derivable from the qualitative nature that the physical mode of presentation represents phenomenal redness as having. Thus not every truth about the qualitative nature that the introspective mode of presentation accurately represents the phenomenal property as having is so derivable. So the physicalist hypothesis is false.¹⁵

One might challenge this version of the knowledge argument at various points. In particular, one might take issue with one or both of the claims about what is intuitive just listed. The one I will dispute is (ii), the claim about the accuracy of introspective representation. I will leave (i) as common ground, and I will continue the discussion with the supposition that (i) is in

¹⁵ On an alternative version of the “old fact/new guise” response, phenomenal modes of presentation should not be taken to represent phenomenal properties as having a qualitative nature at all. Rather, they are just devices for securing reference to phenomenal properties, analogous to demonstratives. There would then be no good reason to think that the physical and phenomenal modes of presentation of phenomenal properties are not co-referential. To my mind, this sort of response to the knowledge argument is weakened in its effectiveness by the plausibility of the claim that phenomenal modes of presentation represent phenomenal properties as having a qualitative nature.

fact true. On (ii), in my view it is an epistemic possibility of a certain sort that introspective modes of presentation do not represent phenomenal properties accurately in the sense just specified, and that these properties might therefore not be as they are introspectively represented. Of the many notions of epistemic possibility, the sense I here have in mind is the usual *possible for all we know*; that is, *possible given what we human beings now know*. (The relevant “we” in this case are perhaps those who have thought carefully about these philosophical issues) For this sense of epistemic possibility, I will use the term ‘*open possibility*’.

My contention, then, is that, given the supposition of (i), it is an open possibility that there is a respect in which introspective modes of presentation do not represent phenomenal properties accurately, and that these properties might therefore not be as they are introspectively represented. For instance, it is an open possibility that when Mary senses the red tomato, her introspective representation of phenomenal-redness presents that property in a *what-it-is-like-to-sense-red* way, while such a qualitative nature is incorrectly attributed to the phenomenal property.

The notion that there might be such a discrepancy between the real qualitative nature of phenomenal properties and the qualitative nature we introspectively represent them as having is consistent with certain claims about the correctness of introspective representation. For example, even if introspective representation inaccurately represents phenomenal properties in the sense just outlined, still it may be that a belief *that I am in* a phenomenal state characterized by a certain phenomenal property, a belief that is formed on the basis of an introspective representation (perhaps a belief that does not feature a linguistic term for the phenomenal state), is generated by a mechanism that is very reliable. So in general there might be no discrepancy

between which phenomenal states I introspectively represent myself as being in and those I am actually in – introspective representation might sort phenomenal states and properties quite accurately -- while at the same time phenomenal properties lack the qualitative nature we introspectively represent them as having.

In this view, a type of representation might successfully secure a referent by, for example, having instances that are caused by this referent, and yet misrepresent this referent by representing it as having a property that it really lacks. Locke's conception of sensory secondary quality representation provides an analogy. He thinks that these representations do indeed secure their referents causally, while they nevertheless misrepresent external objects in a certain respect:

Ideas of primary qualities are resemblances; of secondary, not. From which I think it easy to draw the observation that the ideas of primary qualities are resemblances of them and their patterns do really exist in the bodies themselves, but the ideas produced in us by these secondary qualities have no resemblance of them at all. There is nothing like our ideas existing in the bodies themselves.¹⁶

On a plausible interpretation of this view, our ordinary tactile representations of temperature represent the ambient air, or the icicles above the door, or the coffee one is drinking, as having certain features, while those features are incorrectly attributed to those things. On a warm day, we have a particular sort of tactile temperature representation of the ambient air, which represents the air as having a certain feature – put in Lockean terms, as having a quality that

¹⁶ John Locke, *An Essay Concerning Human Understanding*, (Oxford: Oxford University Press, 1975) II, viii. For a sympathetic exposition of Locke's position on this issue, see Michael Jacovides, "Locke's Resemblance Theses," *The Philosophical Review* 108 (1999), pp. 461-96.

resembles the sensory temperature idea. However, if Locke is right, that “primitive” quality is incorrectly attributed to the air. William Alston might well be endorsing a view of this kind when he says: “when I look at a shirt and take it to be red, when I feel a fabric and recognize it as very smooth, when I hear a bell ringing and recognize it as giving out a typical bell-like sound, I attribute to the perceived objects qualities that they do not, in strictness, bear.”¹⁷

Locke’s contentions about sensory representations of secondary qualities are controversial. Some would dispute the claim that there is any sense in which our ordinary visual color representations generally misrepresent, for the reason that what a type of representation represents is determined solely by the typical cause of its instances. Claims of this last sort have often been disputed by way of devices such as inverted spectrum thought experiments, in which what is represented in the external world is held fixed, while the phenomenal content of the representation varies. Familiarly, there is widespread disagreement about the force of the resulting argument. Nevertheless, I will make use of the secondary quality analogy, assuming Locke’s position that, for example, our ordinary visual color sensations represent the air as having a qualitative feature that is incorrectly attributed to it. A more localized example is that, as Descartes pointed out, from a certain distance we visually represent square towers as round,

¹⁷ William Alston, “Mystical and Perceptual Awareness of God,” in *The Blackwell Guide to Philosophy of Religion*, William E. Mann, ed. (Oxford: Blackwell Publishers, 2004), pp. 198-219, at p. 211. Alston continues: “No doubt, I could, in principle, restrict myself to beliefs that do not suffer falsity in this respect. I could, instead of taking the shirt to be red, take it to have primary qualities of such a sort that when it is seen under these conditions by a human being with normal vision, it will appear to have the color I call red. But that requires considerable reflection of the sort we do not typically engage in when perceiving things.”

while the property of roundness is incorrectly attributed to the tower.¹⁸ Another is that we visually misrepresent the Müller-Lyer pair of lines as having different lengths. It is the open possibility of an analogous disparity between how phenomenal properties are represented introspectively and their real nature that would allow for their being physical despite how they are introspectively represented.

In the case of our visual color representations, the specific causal nature of these representations renders it an open possibility that there is a disparity between how something in the world is represented and how it is qualitatively in itself. By the standard causal theory of external sensory representation, our sensory representations of external objects are entities caused by those objects and distinct from them, which mediate the representational relationship between the subject and the object represented. Due to these mediating sensory representations, there can be a difference between how these objects appear qualitatively by virtue of these representations and how they really are, with the result that these sensory representations are in a respect inaccurate. It is an open possibility that our introspective representation of phenomenal properties is causal in a way analogous to sensory representation of external objects on the usual understanding, whereupon a guarantee of the accuracy of such introspective representation is precluded. Accordingly, it is also an open possibility that our introspective representations of phenomenal properties represent them as having a qualitative nature that they really lack.

Sensory representations of features of external objects can vary as to how accurately they represent those objects. On the Lockean conception, our tactile representations of temperature

¹⁸ René Descartes, *Meditations on First Philosophy*, in *The Philosophical Writings of Descartes*, tr. John Cottingham, Robert Stoothoff, and Dugald Murdoch, v. 2. (Cambridge: Cambridge University Press, 1984); Meditation 6.

represent quite inaccurately. By contrast, our ordinary visual shape representations – which are also causal in nature – would seem to represent the shapes of nearby middle-sized objects more accurately. But it has not been ruled out by anything we now know that our introspective representations of phenomenal states are analogous to our tactile representations of temperature on Locke’s conception; it may be that these introspective representations are quite inaccurate in their representation of these states. In fact, nothing we currently know establishes the degree of accuracy to which our introspective representation of phenomenal states represents those states. We do not know whether, on analogy with the example of tactile temperature representations, introspective representation quite inaccurately represents phenomenal states; or whether it more accurately represents phenomenal states on analogy with some of our visual shape-representations; or whether it represents phenomenal states in a way that guarantees that they are as we so represent them. Each one of these options is an open possibility.

Alternative, non-causal theories of introspective representation are contenders as well. For example, one might, with Franz Brentano, endorse a *self-presentation* view.¹⁹ It claims that a token sensation of green, for example, is on the one hand a sensation of green, but at the same time that very sensation is also *an experience of itself* – or, alternatively expressed, that besides representing to the subject the property of being green, this sensation also simply presents itself to her without the mediation of a (further) representation of it. So, on this view, in one kind of case – when a sensation is an experience of itself -- representation of something occurs without

¹⁹ Franz Brentano, *Psychology from an Empirical Standpoint*, tr. A. C. Rancurello, D. B. Terrell, and L. L. McAlister (London: Routledge and Kegan Paul, 1973), pp. 153-4. Uriah Kriegel develops this position in forthcoming work.

causal mediation. Representation is instead reflexive and non-causal.²⁰ Perhaps the self-presentation view meshes with certain of our ordinary intuitions about our consciousness of sensation. To my mind it is also an open possibility.

Or with David Chalmers one might advocate a *constitution* view for (pure) phenomenal concepts. He says: “one might say very loosely that the referent of the concept is somehow present inside the concept’s sense, in a way much stronger than in the usual cases of ‘direct reference’... in the phenomenal case, the epistemic content itself seems to be constituted by the referent.”²¹ Here again a phenomenal property would be represented without causal mediation. Perhaps this position is not at odds with the self-presentation view, but one might envision it being developed so that it is clearly different.

It may seem strongly intuitive that phenomenal properties are introspectively represented in an intimate way that guarantees that their qualitative nature is represented accurately. The color of a physical object might not be accurately represented by way of our ordinary sensory representations, but how could the qualitative nature of pleasure, or the qualitative nature of

²⁰ Christopher Hill and Brian McLaughlin explain this position as follows:

Sensory states are self-presenting states: we experience them, but we do not have sensory experiences of them. We experience them by *being in* them. Sensory concepts are recognitional concepts: deploying such concepts, we can introspectively recognize when we are in sensory states simply by focusing our attention directly on them. Matters are of course quite different in the case of perceptual and theoretical concepts. An agent’s access to the phenomena that he or she perceives is always indirect: it always occurs via an experience of the perceived phenomena that is not identical with the perceived phenomena, but rather caused by it.

Christopher Hill and Brian McLaughlin, “There Are Fewer Things in Reality than Are Dreamt of in Chalmers’s Philosophy,” *Philosophy and Phenomenological Research* 59 (1999), pp. 445-54, at p. 448.

²¹ David Chalmers, “The Content and Epistemology of Phenomenal Belief,” pp. 13-4.

one's visual sensation of red, not be as it is introspectively represented? But although this sort of discrepancy might be at odds with strong intuitions, still its being an open possibility is forced on us by the prospect that introspection might be causal on analogy with visual color representation.

Moreover, the way we naturally come to think that there are mediating representations in the case of external sensory representation is that here we fairly easily and frequently come to believe that there is a discrepancy between what is represented and its representation. The car appears to have a different color under the sodium vapor lights than it does in natural light, but it is clear that nothing about the car itself has changed, and so we come to believe that there is a discrepancy between the car's real color and the way it is visually represented under the unusual lighting conditions. But for introspection of phenomenal properties, awareness of such discrepancies would not readily arise, supposing they existed. Perhaps they *sometimes* arise:

Christopher Hill cites the following case, presented by Rogers Albritton in seminar:

The case involves a college student who is being initiated into a fraternity. He is shown a razor, and is then blindfolded and told that the razor will be drawn across his throat.

When he feels a sensation he cries out: he believes for a split second that he is in pain.

However, after contemplating the sensation for a moment, he comes to feel that it is actually an experience of some other kind. It is he decides, a sensation of cold. And this belief is confirmed when, a bit later, the blindfold is removed and he is shown that his throat is in contact with an icicle rather than a razor.²²

There are a number of ways to analyze this example, but one possibility is that in his

²² Christopher Hill, *Sensations: A Defense of Type Materialism* (Cambridge: Cambridge University Press, 1991), pp. 128-9.

introspective awareness, the fraternity pledge misrepresents the qualitative features of the sensation of cold he actually has as qualitative features of pain. But this would be a controversial analysis.²³ Here is a possible example of a pain sensation that is introspectively misrepresented as a sensation of cold. My daughter recently required a Novocain shot at the dentist's. Rather than simply showing her the needle in advance, and then giving her the injection, the dentist hid the needle from her, and told her that he would be dropping bits of cold water into her mouth. She didn't flinch. When I asked her afterward whether the experience was unpleasant, she said that she didn't like the drops of water much, but they didn't hurt. Here we might want to say that the dentist's suggestion, together with his hiding the needle, kept her from introspectively representing the qualitative features of the pain state she was actually in as qualitative features of pain, while instead she misrepresented those features as qualitative features of a sensation of cold. This, again, is a controversial analysis. But my point here is that if such examples of misrepresentation occur at all, it is only infrequently. They are too unusual to give rise to a vivid sense that for introspection of phenomenal properties, there is a discrepancy between what is represented and its representation.

In addition, in the external case, we have readily available ways of checking the object

²³ Hill takes this example to provide evidence that we make *errors of judgment* in our introspection-based beliefs about sensation, where errors of judgment "are usually due either to some form of inattention or to the influence of expectation upon judgment." He differentiates between errors of judgment and *errors of ignorance*, which occur "when beliefs are based on appearances that fail to do justice to the entities to which the beliefs refer." Hill claims "we are perforce innocent of committing errors of ignorance in forming beliefs about our own sensations;" *Sensations*, p. 127-8. The open possibility I am envisioning would have us making errors of ignorance in our introspection-based beliefs about phenomenal properties, since such beliefs would be based on appearances that fail to do justice to the real qualitative nature of those properties.

represented that are independent of the representation under scrutiny, while in the introspective case such a capacity is at best very limited. One might have a closer look at Descartes's tower in order to check whether one's visual representation of its shape as round was accurate, or measure the Müller-Lyer lines to determine whether one's visual representation of them as having different lengths was correct. But analogous ways of checking introspected phenomenal properties are not similarly available, if we have them at all.

These observations help explain our resistance to the idea that introspection of phenomenal properties features mediating representations. Given that introspective awareness of a discrepancy between the phenomenal property represented and its representation would seldom arise, and given the scarcity of means of checking the accuracy of such representations, there would be little difference between an introspective experience in which we represented phenomenal properties causally, and one in which phenomenal properties (or the sensations or states of which they are properties) were self-presenting, or one for which the constitution view held. Hence, introspective experience, all by itself, would not adjudicate the controversy as to how phenomenal properties are represented.²⁴

While the self-presentation and constitution views are candidates for the correct account of introspective phenomenal representation, a causal theory is also contender. Thus, while the

²⁴ Stephen Wykstra proposes the following plausible *condition of reasonable epistemic access*: "On the basis of cognized situation *s*, human *H* is entitled to claim 'It appears that *p*' only if it is reasonable to believe that, given her cognitive faculties and the use she has made of them, if *p* were not the case, *s* would likely be different than it is in some way discernible by her;" ("The Humean Obstacle to Evidential Arguments from Suffering; On Avoiding the Evils of 'Appearance'," *International Journal for Philosophy of Religion* 16 (1984), pp. 73-93; reprinted in *The Problem of Evil*, M. M. Adams and R. M. Adams, eds., pp. 138-60, at p. 152). By Wykstra's criterion, we would not be justified in claiming that it appears that introspective representation of phenomenal properties is non-causal.

self-presentation and constitution views are open possibilities, a causal account of introspective phenomenal representation is a serious open possibility as well. It is then also a serious open possibility that introspection represents the qualitative nature of phenomenal properties inaccurately.

It is worth noting that the self-presentation view of phenomenal representations does not obviously preclude such qualitative inaccuracy. Self-presenting sentences can misrepresent in some respect (while being accurate in another) -- ‘this German sentence has six words,’ is a case in point.²⁵ Perhaps, then, nothing we know rules out the possibility that self-representing phenomenal states be qualitatively inaccurate in the sense required for the argument. It may also be that the constitution view does not preclude qualitative inaccuracy – we would need to be told more about how it works. Still, I suspect that the stronger case for my position can be made by analogy with secondary-quality representation, and here it is reasonable to believe that qualitative inaccuracy is due to causal mediation, and not to the sort of problem that arises in the case of misrepresenting self-presenting sentences. Nonetheless, if qualitative inaccuracy for introspective representations of phenomenal properties is an open possibility even if phenomenal states are self-presenting, then my case will be stronger.

It should be noted that in the proposed open possibility, the qualitative inaccuracy for our introspective representations of phenomenal properties is universal -- it is a feature of all such

²⁵ Mark Moyer and Brian Weatherson each made this point about self-representing sentences and suggested that the possibility of misrepresenting self-representation would strengthen the argument. Louis deRosset provided the example of a self-presenting sentence that is accurate in one respect and inaccurate in another.

human introspective representations; and it is extensive -- there is little if anything in common between the real qualitative nature of phenomenal properties and how they are represented introspectively. In these respects, the inaccuracy at issue differs from that of the sort we sometimes make when we visually represent the lengths of pairs of lines, or when visually represent the shapes of objects from a significant distance.²⁶ One might contend that the universality and extensiveness of the proposed inaccuracy provides reason to believe that the open possibility under consideration is very unlikely to be actual. However, on the Lockean theory of our sensory representations of secondary qualities, which is not an implausible theory, the inaccuracy of these representations is similarly universal and extensive. To my mind, this analogy yields significant reason to believe that the proposed open possibility is not unlikely to be actual.

These reflections suggest the following reply to the knowledge argument. While Mary is in the room, she does not represent a phenomenal-red sensation in the characteristic introspective way. But it is a serious open possibility that by virtue of her physical knowledge she nevertheless accurately represents the complete real nature of this phenomenal state and its properties. For the qualitative nature of the property of phenomenal redness, in particular, might not be as it is introspectively represented. Instead, it might be accurately represented by virtue of Mary's physical knowledge. Accordingly, from her physical knowledge she might then be able to derive every truth about the real nature of the phenomenal state, despite the fact that this physical

²⁶ Thanks to Louis deRosset for this point.

knowledge gives her no access to the phenomenal state as it is introspectively represented.²⁷

On this open possibility, how exactly should we describe what happens when Mary leaves the room and sees the red tomato? We imagine her coming to believe that

(T) phenomenal redness has qualitative nature Q.

Consider first the initially plausible proposal (i) that in (T), the term ‘Q’ refers to a property accurately represented by the what-it-is-like-to-sense-red mode of presentation. On our open possibility, there is no such property, and so that what she comes to believe will be false. Then since she merely acquires a false belief, she fails to gain knowledge. Next consider the initially less plausible proposal (ii) that the term ‘Q’ in (T) refers to a physical property that appears to Mary in the what-it-is-like-to-sense-red way, but whose qualitative nature is not accurately represented by this mode of presentation. Under this interpretation we can suppose that (T) is true, but then since she while in the room she already knew the fact expressed by (T), she does not gain knowledge.

²⁷ Chalmers (in “Phenomenal Concepts and the Explanatory Gap,” *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Torin Alter and Sven Walter, eds., (Oxford University Press, 2006), pp.) points out that on the sort of view advocated by Loar, Levine, and others -- the “phenomenal concept strategy” -- it is maintained that zombies are ideally, positively, primarily conceivable, while our having phenomenal concepts has a physical explanation. These are two key features of what Chalmers calls Type-B materialism, a widely held view. He argues that this position is unstable. I suspect that he is right about this, and that in the last analysis, physicalism requires denying the ideal, positive, primary conceivability of zombies, for then it would avoid the tension between affirming the conceivability of zombies, which has the consequence that phenomenal properties are in the crucial sense not physically explainable, and claiming that our having phenomenal concepts is physically explainable. Another key feature of the phenomenal concept strategy is its claim that ‘P → Q’ is a posteriori, and Daniel Stoljar (“Physicalism and Phenomenal Concepts”) argues that it cannot adequately explain how this can be so. The response I’ve developed here does not require that this conditional is a posteriori, and thus it avoids the issue Stoljar highlights.

Thus, speaking within the scope of the open possibility under consideration, what Mary comes to believe when she leaves the room would be false on the initially plausible interpretation (i), and while on the initially less plausible proposal (ii) it would be true, she would not gain knowledge. The intuition that fuels the knowledge argument is that (T) is true, and thus what Mary comes to believe is true, and as a result she gains knowledge. How might we then explain away the strength of this intuition that fuels the knowledge argument? In believing that she gains knowledge, and that (T) is true, Mary is *mistaking appearance for reality in a way that is extremely natural*. In particular, she is assuming that the real qualitative nature of the phenomenal property she is representing *is* as it is presented in introspective experience. The existence of a disparity between appearance and reality at these points is, as I have argued, a supposition that would not ordinarily arise. But nonetheless, there might be such a disparity. It might well be that the real qualitative nature of these properties is not as it introspectively appears, and if it isn't, the intuition that fuels the knowledge argument might well be false. Moreover, if one holds that there is a disparity between how colors appear to visual representation and how they really are, as many do, then it is not obvious that one can legitimately reject out of hand the possibility of an analogous disparity between how phenomenal properties appear to introspection and how they really are.²⁸

²⁸ David Chalmers's zombie argument can be challenged in a similar way (David Chalmers, *The Conscious Mind* (Oxford: Oxford University Press, 1996); "Consciousness and its Place in Nature," *Blackwell Guide to the Philosophy of Mind* (Oxford: Blackwell, 2002); reprinted in David Chalmers, ed. *Philosophy of Mind: Classical and Contemporary Readings* (Oxford: Oxford University Press, 2002), pp. 247-72.) Consider the first premise of the zombie argument. Let 'P' be a statement that details the complete physical truth about the actual world, and 'Q' be an arbitrarily selected actual phenomenal truth. The first premise is then: 'P and ~ Q' is ideally, positively, primarily conceivable. This premise requires the crucial assumption that we have representations that accurately represent the qualitative nature of phenomenal properties and the

One might think that despite these considerations, it remains implausible to claim that the qualitative inaccuracy for our introspective representations of phenomenal properties is as universal and extensive as would yield a promising materialist response to the knowledge argument, and this should provide resistance to belief that this claim is true. But it may be that all of the developed positions on the metaphysics of consciousness have implausible features that should make for at least some resistance to belief. As a case in point, Karen Bennett argues that the traditional dualist position has such an implausibility: it accepts that there exist a fairly large number of psychophysical laws that are brute in the sense that there is no explanation as to why they hold, or for which the explanation we can envision is arbitrary divine preference.²⁹ (Locke suggests the divine preference explanation, and Adams endorses it.³⁰) With this in mind, should one be less resistant to the traditional dualist view than to the qualitative inaccuracy claim?

states that have them. More specifically, it requires that if it is not the case that a state has a property that is accurately represented by an introspective mode of presentation of the what-it-is-like-to-sense-x variety, it is not the case that the phenomenal property represented by this mode of presentation is instantiated by the state. For example, consider the phenomenal concept 'R', a concept of phenomenal property R, which represents R by way of the introspective mode of presentation *what-it-is-like-to-sense-red*. The argument demands that if it is not the case that a state has a property whose qualitative nature is accurately represented by the what-it-is-like-to-sense-red introspective mode of presentation, then that state does not instantiate phenomenal property R. But for this last conclusion to be established, it would have to be shown that it is not an open possibility for the phenomenal concept 'R' to be qualitatively inaccurate. For if this representation of R is indeed qualitatively inaccurate in this respect, then the fact that R *as represented in the what-it-is-like-to-sense-red way* is not derivable from 'P' fails to show that a description of the *real nature* of R is not derivable from 'P.' Thus it also does not show that 'Q', our selected truth about R, is not derivable from 'P'. Then, what we were thinking of as the zombie-world might not be one in which 'Q' is false after all.

²⁹ Karen Bennett, "Why I Am Not a Dualist," manuscript.

³⁰ John Locke, *Essay Concerning Human Understanding* IV, iii, 6, 28-9; Robert Adams, "Flavors, Colors and God," pp. 241-51.

Alternative physicalist hypotheses also have elements that are to at least some degree implausible, as their opponents, such as Chalmers, have contended.³¹ Should one be less resistant to these physicalist views than to the qualitative inaccuracy claim? There are a number of central philosophical issues for which all defended positions are in some key respect implausible – free will and moral responsibility is a case in point. For such issues, it is not sufficient to dislodge a position to show that is in some way implausible, and the metaphysics of consciousness might well be such an issue.

2. Hasn't the problem for physicalism been shifted to modes of presentation?

However, it may now seem that the problem for a physicalist explanation of consciousness has shifted from accounting for phenomenal states and their properties to accounting for their introspective modes of presentation. Supposing that the way phenomenal states are represented introspectively might be inaccurate, and that Mary can derive every truth about the real nature of phenomenal states from her complete physical knowledge, the pressing issue is now to assess whether these introspective modes of presentation, or perhaps more precisely, states featuring these modes of presentation, could have a physical account.³² In fact, Chalmers develops this point as an objection to the old fact/new guise strategy. He contends that even if what Mary gains when she leaves the room

is only knowledge of an old fact under a different mode of presentation – then there must be some truly novel fact that she gains knowledge of. In particular, she must come to

³¹ See, for example, David Chalmers, “Consciousness and its Place in Nature.”

³² I consider this objection in “Bats, Brain Scientists, and the Limitations of Introspection,” *Philosophy and Phenomenological Research* 54 (1994), pp. 315-29, at pp. 323-6.

know a new fact involving that mode of presentation. Given that she already knew all the physical facts, it follows that materialism is false. The physical facts are in no sense exhaustive.³³

Torin Alter raises a similar objection to my earlier account:

How colour sensations appear from the first-person perspective is itself a fact about them. Therefore, if when Mary is released she learns how they appear from the first-person perspective, then she learns a new fact about them. This is true regardless of whether this appearance accurately reflects the way they really are.³⁴

So would Mary, by virtue of her physical knowledge, be able to derive every truth about the introspective mode of presentation of her phenomenal-red sensation – call it MP_R , or every truth about representational states that have MP_R as a component? In response, there is no less reason to think that a causal theory is true for introspective representations of introspective modes of presentation, or for introspective representations of states that have introspective modes of presentation as a component, than it is for phenomenal states themselves. Consequently, it is also a serious open possibility that our introspective representation of a state that has MP_R as a component is qualitatively inaccurate. Then, despite how it may be introspectively represented, it might be that Mary can derive every truth about the real nature of a state that has MP_R as a component from her physical knowledge. So even though pre-emergence Mary will not have an introspective representation of this state, it is an open possibility that while she is in the room she

³³ David Chalmers, *The Conscious Mind*, p. 142.

³⁴ Torin Alter, “Mary’s New Perspective,” *Australasian Journal of Philosophy* 73 (1995), pp. 582-4.

can come to know every truth about it. Furthermore, a reply of this kind can be made to count against any iteration of this sort of objection.³⁵

The success of the knowledge argument depends on there being phenomenal states or aspects of phenomenal states whose qualitative nature is as it is introspectively represented. Alter and Chalmers are right to argue that the “old fact/new guise” response to the knowledge argument typically transfers the physicalism-challenging feature from the phenomenal state to the introspective mode of presentation. However, the basic “old fact/new guise” move can be reiterated for introspective modes of presentation. Notice that the view that results from this is longer best classified in the “old fact/new guise” category. For Mary knew everything there is to know about MP_R , and states featuring it, prior to emerging from the room, and thus there is a clear sense in which the guise is not new. At the same time, prior to emerging from the room, Mary had never represented a phenomenal property by means of this mode of presentation, and hence one might say that her deployment of MP_R is new.

3. Phenomenal concepts and conceptual analysis.

To all of this one might object that conceptual analysis of our phenomenal concepts reveals that they apply correctly to properties whose qualitative nature is accurately represented by introspection, and that it is ruled out conceptually that they correctly apply to properties whose qualitative nature is not accurately represented in this way. Chalmers suggests an idea of this sort when he specifies that the referent of a pure phenomenal concept is present inside the concept's

³⁵ One might object that this move generates a vicious regress; for a discussion of this objection see my “Bats, Brain Scientists, and the Limitations of Introspection,” pp. 325-7.

sense, and its content is constituted by the referent.³⁶ Some of his physicalist opponents concur. Brian Loar, for example, argues that phenomenal concepts *express* the very properties they pick out. In his framework, a concept expresses its reference-fixer, and thus he is contending that reference-fixers of phenomenal concepts are just the properties they pick out. Moreover, he claims that:

Phenomenal concepts pick out certain properties directly. They do not pick out those properties via a contingent mode of presentation, in the manner say of visual recognitional concepts, which connect one to some external kind by way of a visual experience. It could then seem, I suppose, that phenomenal concepts conceive their referents *as they are in themselves*.

Loar is plausibly interpreted as endorsing the claim that phenomenal concepts accurately represent the qualitative nature of the properties to which they apply.

To evaluate this objection, we first need to be clear about what it is that conceptual analysis reveals. An attractive proposal that derives from Hilary Putnam is that the structure of certain concepts is a conjunction of conditionals. The antecedents of the conditionals specify coherent scenarios considered as actual – that is, considered as if they were the way things actually turned out, and the consequents indicate what the concept in question then correctly applies to. Which conjunct is *operative* depends on the way the world actually is, since which

³⁶ David Chalmers, “The Content and Epistemology of Phenomenal Belief,” pp. 13-4. Brian Loar, “David Chalmers’s *The Conscious Mind*,” *Philosophy and Phenomenological Research* 59 (1999), p. 471; cf. Brian Loar, “Phenomenal States,” in *The Nature of Consciousness: Philosophical Debates*, Ned Block, Owen Flanagan, Guven Güzeldere (Cambridge: MIT Press, 1997).

conjunct is operative is a matter of which conjunct's antecedent is true.³⁷ This structure is discerned by reflection on possible cases -- (Jackson makes an impressive case that the sort of reflection on possible cases that we see in Putnam's work might be thought of as conceptual analysis).³⁸

For example, given that all of our samples of the watery stuff in our environment are constituted of H₂O, and this chemical property explains the properties we associate with water, our concept 'water' correctly applies (just) to H₂O, and water = H₂O. But suppose that it had turned out that this watery stuff, like our samples of jade, had two distinct kinds of composition, each at least fairly common. Then claiming that 'water' correctly applies only to H₂O would have been implausible, and, like jade, it would have turned out that water was a disjunctive kind.³⁹ Or further, imagine that it instead turned out that this watery stuff had many distinct constitutions with no salient similarities among their intrinsic features, while each sample nevertheless exemplified a well-behaved functional characterization. Then, like 'catalyst' and 'enzyme,' we

³⁷ Hilary Putnam, "The Meaning of 'Meaning'", in his *Philosophical Papers*, Volume 2, (Cambridge: Cambridge University Press, 1975), pp. 240-1. This idea has been endorsed and developed by George Bealer, "Modal Epistemology and the Rationalist Renaissance," in *Conceivability and Possibility*, Tamar Szabó Gendler and John Hawthorne, eds. (Oxford: Oxford University Press, 2002), pp. 77-125, at p. 109; Ned Block and Robert Stalnaker, "Conceptual Analysis, Dualism, and the Explanatory Gap," *The Philosophical Review* 108 (1999), pp 1-46, at p. 36; and Jackson and Chalmers would not dissent from this line of thought, David Chalmers and Frank Jackson, "Conceptual Analysis and Reductive Explanation," *Philosophical Review* 110 (2001), pp. 315-61, esp. pp. 322, 340-1.

³⁸ Frank Jackson, *From Metaphysics to Ethics* (Oxford: Oxford University Press, 1998), pp. 28-86.

³⁹ The 'would have turned out that S, had it turned out that W' locution derives from Stephen Yablo, "Shoulda, Woulda, Coulda," in *Conceivability and Possibility*, Gendler and Hawthorne, eds., pp. 441-92, at p. 454.

might have correctly counted water as a functional kind. Or suppose it turned out that Berkeley's view of the universe was correct, and that water was composed just of sensations directly produced in our minds by God. Then we might have classified water as an appearance kind, so that 'water' applied correctly to anything that appeared in one particular way under certain conditions, and in a different particular ways under other conditions. If it seems strange that this last scenario has a place in the conceptual analysis of 'water,' it is important to keep in mind that Berkeley's idealist world is not ruled out a priori by conceptual analysis, and that the complete conceptual analysis of 'water' must specify its correct application conditions in any such world, or coherent scenario, considered as actual.⁴⁰

On this proposal, conceptual analysis reveals that our concept 'water' has a structure something like:

If a scenario is actual in which the watery stuff in the environment has a unique sort of composition, then the concept 'water' correctly applies to a unique compositional stuff;⁴¹

and

if a scenario is actual in which the watery stuff has a small number of sorts of composition, then the concept 'water' correctly applies to a disjunctive compositional stuff;

⁴⁰ David Braddon-Mitchell makes a related point in "Qualia and Analytical Conditionals," *Journal of Philosophy* 100 (2003), pp. 111-35, at p. 115. Braddon-Mitchell also suggests that the analysis of some concepts might be a conjunction of conditionals.

⁴¹ The conditionals might also be formulated non-metacognitively, for example: If a scenario is actual in which the watery stuff in the environment has a unique sort of composition, then water is a unique compositional stuff.

and

if a scenario is actual in which the watery stuff has many sorts of composition, and in which there are no salient similarities among the intrinsic properties of these compositions, while each sample of the watery stuff exemplifies a well-behaved functional characterization, then the concept 'water' correctly applies to a functional kind,

and

if a scenario is actual in which each instance of the watery stuff is a collection of sensations produced directly in minds by God, then the concept 'water' correctly applies to an appearance kind...

Let us call conjunctions of conditionals of this variety *Putnam-conjunctions*. For certain concepts, the plausibility of this picture serves as a corrective to the idea that conceptual analysis alone can determine, in effect, that a specific conditional is operative. For example, it is sometimes assumed that conceptual analysis alone shows that 'water' refers to a unique compositional stuff. But this would then not be so -- in this case, conceptual analysis would reveal only a conjunction of conditionals. Which of the conjuncts is operative would then be settled by the actual world, and we would know which conjunct is operative only by our investigation of the actual world. This model allows a concept to remain the same through changes in our scientific theories about what it correctly applies to, or (more salient for present purposes) through a more rudimentary change from a situation in which we rely only on the manifest image for its conditions of correct application to one in which we are informed by a scientific theory. For the model allows that the concept remains the same through such changes,

while what is considered to be the operative Putnam-conjunct varies.

Returning to our secondary quality analogy, when examining the nature of color concepts, one might claim that conceptual analysis reveals that

C1. Redness is the property of objects that is the normal cause of their looking red (where ‘the normal cause of their looking red’ functions merely as a reference-fixer).

But what if it turns out there are many different sorts of causes of looking red, and there are no salient similarities among the intrinsic properties of these causes? C1 might then predict that then there is no such property as redness, or at least that redness is not instantiated -- if wildly disjunctive properties are excluded, for example. However, the same might then need to be said about any proposed response-dependent property whose categorical basis was wildly disjunctive, which would be implausible. Then it might turn out that:

C2. Redness comprises whatever properties cause (or could cause) instances of looking red.

Or suppose that because Berkeley’s theory turned out to be true, God was the normal cause of objects’ looking red. Would we then say that God is red? More likely, redness would then be an appearance property. So then, while conceptual analysis of color concepts might initially seem to reveal something like C1, a more thorough analysis would yield a complex Putnam-conjunction.

There is a moral here for the analysis of phenomenal concepts. Consider the claim that conceptual analysis alone reveals that phenomenal concepts refer to properties that resemble our introspective representations of them, so that

P3. Phenomenal redness is the property that resembles the introspective representation of it.

But by analogy, consider an Aristotelian who holds that conceptual analysis reveals that

C3. Redness is the property of objects that resembles sensations of red.

Suppose he is confronted with a convincing scientific demonstration that physical objects have no such properties. He might conclude that redness is not instantiated in the physical world -- that the concept 'red' does not correctly apply to anything in the physical world, as Galileo did.⁴²

But almost everyone today believes that a response of this sort is mistaken, and that (what is actually) a different conjunct of our concept of red, such as the one that reflects C1, would be operative. Similarly, investigation might lead us to think that the operative conjunct of a phenomenal concept is not the one that has it applying correctly to a property that is accurately represented introspectively. Suppose it turns out that there are no instantiated properties accurately represented by introspective phenomenal representations. One might conclude that phenomenal concepts fail to apply to any instantiated properties. However, as in the case of color concepts, a radical conclusion of this sort is not clearly forced on us. Rather, it might well be that there are alternatives reflected in other conjuncts in the analysis of phenomenal concepts.

It might be objected that while it is possible to devise phenomenal concepts whose analysis is complex in this way, still our ordinary phenomenal concepts are simple in the sense of being non-conjunctive, and have something like P3 as an analysis, so whether there are phenomenal properties on the ordinary understanding depends on whether there are properties that fit something like P3. But one might envision Aristotelians about color having made an analogous claim: "One might devise color concepts whose analysis is a complex conjunction of

⁴² Galileo Galilei, *The Assayer*, in *Discoveries and Opinions of Galileo*, tr. Stillman Drake, (New York: Doubleday Anchor, 1957), pp. 217-80, at pp. 274-7.

conditionals, but ordinary color concepts are simple, and are to be analyzed on the order of C3.” But, as history has shown, the initial attractiveness of C3, and its resilience, is insufficient to show that it provides the complete and correct analysis of our concept of red. The case of phenomenal redness is, I suggest, parallel. The initial attractiveness of something like P3, and its resilience, is insufficient to show that it is the complete and correct analysis of our concept of phenomenal redness. Rather, it is an open possibility that the analysis of this concept reveals a complex Putnam-conjunction, and that the operative conjunct renders true a different sort of characterization, such as:

P1. Phenomenal redness is the property that is the normal cause of introspective representations of phenomenal redness (where ‘the normal cause of introspective representations of phenomenal redness’ functions merely as a reference-fixer).

or

P2. Phenomenal redness comprises whatever properties cause (or could cause) instances of the introspective representation of phenomenal redness.

Even if typical current theories of phenomenal concepts attribute qualitative accuracy to them, dispensing with those theories might well not be ruled out by the nature of the concepts themselves, and may be welcome. The Aristotelians maintained that our concept of temperature is a sensory concept, and that it generally represents the qualitative nature of temperature accurately. By contrast, Locke argued that our temperature concept is a secondary quality idea – a type of response-dependent concept -- and that it or its sensory content represents the qualitative nature of temperature inaccurately.⁴³ A Kripkean understanding has it that our

⁴³ John Locke, *An Essay Concerning Human Understanding*, II, viii, 15.

concept of temperature, like our concept of water, is not a secondary quality concept, but that it is rather a natural kind concept, and again – at least given Lockean intuitions – that our tactile representations of temperature are qualitatively inaccurate. Plausibly, this evolution of theory about our concept of temperature amounts to progress – (perhaps) we have a better understanding of which conjunct of the Putnam conjunction for ‘temperature’ is actually operative, than the Aristotelians did. Similarly, even though it may currently be compelling to theorize that our phenomenal concepts are qualitatively accurate, we may be led to conclude otherwise. It is an open possibility that a causal account of phenomenal representation is true, and this gives rise to the open possibility that phenomenal concepts are qualitatively inaccurate. Then we may find that for each type of introspective phenomenal representation, there is a single type of physical property that is its normal underlying cause, whereupon we might conclude that phenomenal concepts correctly apply to such underlying physical causes. Or else we may find that there are many very different sorts of properties that can cause instances of a single type of phenomenal property, and this might have us come to believe that phenomenal concepts correctly apply to any such properties. The conceptual analysis of ‘phenomenal redness’ might well allow for such alternatives, despite what we may have thought. If one wants to deny this, and claim instead that by conceptual analysis it can be shown that just

P3. Phenomenal redness is the property that resembles the introspective representation of it.

is generally representative of the analysis of the relevant sort of phenomenal concepts, it seems to me that one would need to develop more thoroughly a theory of such concepts (such as the self-presenting or constitution views) that would indicate how it might be that this claim is clearly

true.

4. Edenic and ordinary phenomenal content

The analogy with secondary quality representation can be developed further to explain, on the open possibility under consideration, the strength of the intuition that Mary learns something new upon seeing the tomato. Consider Chalmers's view of the content of phenomenal color representation. He argues that the account of such content that is most adequate to the phenomenology of color perception is primitivism (of which the Aristotelian position is a variety):

The view of content that most directly mirrors the phenomenology of color experience is primitivism. Phenomenologically, it seems to us as if visual experience presents simple intrinsic qualities of objects in the world, spread out over the surface of the object. When I have a phenomenally red experience of an object, the object seems to be simply, primitively, *red*. The apparent redness does not seem to be a microphysical property, or a mental property, or a disposition, or an unspecified property that plays an appropriate causal role. Rather it seems to be a simple qualitative property, with a distinctive sensuous nature. We might call this property perfect redness: the sort of property that may have been instantiated in Eden.⁴⁴

Rather than characterizing primitive properties by way of resemblance to sensations, as Locke

⁴⁴ David Chalmers, "Perception and the Fall from Eden," Tamar Szabó Gendler and John Hawthorne, eds., *Perceptual Experience* (Oxford: Oxford University Press, 2006), pp. 49–125, at p. 66.

does, Chalmers opts characterizing them as properties that are as they seem to sensory experience. To sensory experience these properties appear simple in the sense of not having an internal causal or dispositional structure, and in the sense of not being composed, for example, of microphysical particles. They also appear to be non-mental properties. In addition, our experience of them is not as unspecified properties – we might say that we experience them as having a specific and determinate nature.⁴⁵ The content of a phenomenal color representation associated with these primitive properties Chalmers calls its *Edenic content*.⁴⁶

But Chalmers thinks that science and philosophical reflection provide us with good reasons to believe that there is no instantiated property to which this Edenic content correctly applies – there are no instantiated primitive color properties. However, this does not mean that there are no colors. For there is a veridical content of phenomenal color representation that well-enough *matches* its perfect content – which Chalmers calls its *ordinary content*. Edenic content functions as a kind of regulative ideal in determining the ordinary content of our color experiences -- it is the standard that matching ordinary content must most closely approximate -- but its being merely a regulative ideal allows for matching ordinary content that is veridical.⁴⁷

Notice that this account seems to commit Chalmers to the qualitative inaccuracy of visual color

⁴⁵ In the Garden of Eden, Chalmers specifies, “we had unmediated contact with the world. We were directly acquainted with objects in the world and with their properties. Objects were simply presented to us without causal mediation, and properties were revealed to us in their true intrinsic glory”; “Perception and the Fall from Eden,” p. 48. Chalmers is specifying an ideal here; he does not deny that primitivism about visual color representation can accommodate a causal theory of such representation, as in the Aristotelian view.

⁴⁶ David Chalmers, “Perception and the Fall from Eden,” pp. 69-71.

⁴⁷ David Chalmers, “Perception and the Fall from Eden,” pp. 69-84.

perception. Such perception represents physical objects as primitively colored, but they are not primitively colored, while at the same time they are colored. Visual color perception sorts colors quite correctly, but represents something else about them inaccurately, and the only candidate for what is inaccurately represented would seem to be the color's qualitative nature.

But notice that a story parallel to Chalmers's account of the content of color representation can be given for our introspective representations of phenomenal properties. When I have an introspective representation of phenomenal redness, what I apprehend seems to be simply, primitively, phenomenally red. Phenomenal redness seems to be a simple qualitative property, with a distinctive sensuous nature. We might call this property primitive phenomenal redness, and the content of introspective phenomenal redness associated with this property its Edenic content. But given the open possibility that our introspective representations of phenomenal properties are in actual fact causal and mediated, it may be that these representations are also qualitatively inaccurate, and that their Edenic content correctly applies to no instantiated properties. Still, there might be an ordinary content of these representations that matches their Edenic content closely enough, with the consequence that there are instantiated phenomenal properties to which this matching content correctly applies. These properties might be physical properties, such as the physical property that is the normal cause of introspective representations of phenomenal redness.

Consider two possible proposals for the ordinary content of representations of phenomenal redness (derived from P1 and P2 above):

OC_i: an ordinary content that correctly applies to the property that is the normal cause of introspective representations of phenomenal redness (where 'the normal cause of

introspective representations of phenomenal redness' functions merely as a reference-fixer).

OCii: an ordinary content that correctly applies to whatever properties cause (or could cause) instances of the introspective representation of phenomenal redness.

On OCi, is nomologically possible for a state to be introspected as phenomenal redness while phenomenal redness is not then instantiated, since a property that on some occasion causes the introspective representation of phenomenal redness might not be the property that is its normal cause. Accordingly, a characteristic of OCii that might argue in favor of its being the closest match to the regulative ideal is that it would make it impossible for a state to be introspected as phenomenal redness while phenomenal redness is not instantiated -- so that if a state seems conscious in the what-it-is-like-to-see-red way, it will be conscious in this way. As applied to pain, for example, this might count in favor of OCii, since it is at least initially strongly unintuitive for many that a state be introspectively represented as pain and not be pain. On the other hand, it may count in favor of OCi that it would allow our classification of phenomenal properties to cut nature at its causal joints, after the manner of Kripkean natural kind terms or concepts, while OCii is not designed to do so. I favor OCi. As for its unintuitive consequence, I suspect that it might well be that a state be introspectively represented as pain and not be pain. Consider the reverse of the Novocain example discussed earlier; someone says that he is going to inject a large needle into your mouth, but instead administers drops of cold water. If introspectively, pain can be mistaken for the sensation of drops of cold water, then it would seem that, introspectively, the sensation of drops of cold water could be mistaken for pain. So then a

state might be introspectively represented as being pain, but not be pain.⁴⁸

Given an Edenic content interpretation,

(T) phenomenal redness has qualitative nature Q,

and supposing (T) is true, Mary would learn something new when she leaves the room and sees the tomato. But on the open possibility we are considering, on an Edenic content interpretation, (T) is in fact false. So it would be false that Mary learns something new. On several ordinary content interpretations (for example P1 and P2), it turns out that (T) is derivable from what pre-emergence Mary knows – there would then be no less reason to believe that (T) is derivable from this information than there is to think that ‘more than half of the earth’s surface is covered with water’ is so derivable. Then again Mary would not learn anything new when she sees the tomato, but for a different reason – she already knew (T) when she was in the room. Thus, on our open possibility, for both Edenic and ordinary content interpretations of (T), it is false that Mary learns something new when she leaves the room, and the knowledge argument fails.

5. The explanatory gap and eliminativism.

Chalmers contends that several commentators who have attempted to undermine the knowledge argument (and the zombie argument) by the “old fact/new guise” response have failed to show how it might be that the distinct modes of presentation, physical and phenomenal, might refer to the same thing. This issue is especially pressing for Loar, who holds that a phenomenal concept expresses the phenomenal property to which it refers, while physicalism is true. On

⁴⁸ Thanks to Kati Balog for raising the objection that occasioned these thoughts.

Chalmers's reading, Loar in fact maintains "that phenomenal and physical concepts (i) are cognitively distinct, and (ii) both express the property they refer to."⁴⁹ Chalmers argues that if both (i) and (ii) are accepted, nothing can justify the claim that the phenomenal and physical concepts co-refer. Or perhaps it is actually impossible that (i) and (ii) are both satisfied, for the reason that the phenomenal concept correctly applies to a primitive phenomenal property, while this is not the property the physical concept expresses. However, if phenomenal concepts are qualitatively inaccurate, Chalmers's explanatory burden – which is part of the burden of the explanatory gap – can be discharged. For then it need no longer be explained how a qualitative nature that resembles the what-it-is-like-to-sense-red mode of presentation can *correctly* be attributed to a phenomenal property, while that property is at the same time physical, and a description of its real qualitative nature is represented by or derivable from 'P'.

With regard to this explanatory gap, Adams argues that materialism has no adequate response to the demand to explain why particular kinds of phenomenal properties are correlated with particular kinds of physical properties:

For suppose a materialist claims that [physical property] *R* and the phenomenal appearance of red are one and the same property of brains, identified as *R* on the basis of its place in the physical system, and as the appearance of red on the basis of the way it seems to us when our brains have it. We can still ask why *R* seems to us the way it does, rather than the way *Y* (the physical brain state which "is" the appearance of yellow) does.

⁴⁹ David Chalmers, "Materialism and the Metaphysics of Modality," p. 487-8.

This is quite recognizably our original question, and it remains unanswered.⁵⁰

Adams's demand is for a contrastive explanation: why does physical property *R* seem the way it does, and not the way physical property *Y* seems? Supposing that different brain states appear in different ways to introspection, the physicalist needs to explain why any one brain state appears to introspection in one way, and not in some other way, for example in the way some other brain state appears to introspection. Adams believes that the physicalist has no adequate response to this demand.

But we can reply: it is an open possibility that there is a discrepancy between the real nature of the property *how R seems* and the qualitative nature we introspectively represent this property as having. It is then an open possibility that *how R seems* is a straightforwardly physical property, call it *RS*, despite how we introspectively represent it. The same can be said of *how Y seems* – it might be a straightforwardly physical property – call it *YS*, despite the qualitative nature we introspect it as having. If this open possibility is actual, then the physicalist can meet the demand for contrastive explanation, which might then be formulated as: why does physical property *R* cause physical property *RS* and not physical property *RY*? We're assuming that Mary, while she is in the room, has mastered purely physical explanations of this sort. So on the open possibility under consideration, the physicalist can meet Adams's demand for an explanation.⁵¹

⁵⁰ Robert Adams, "Flavors, Colors, and God," p. 259.

⁵¹ As mentioned in note 7, Stoljar's version of the knowledge argument specifies that pre-emergence Mary possesses all the phenomenal concepts, while she nevertheless lacks knowledge of how correct applications of phenomenal concepts are correlated with physical states ("Physicalism and Phenomenal Concepts"). The anti-physicalist might then contend that while she is in the room, Mary would not be able to produce contrastive explanations of the sort Adams

Adams further contends that a materialistic explanation of correlations between physical and phenomenal properties would require a materialism of a radical sort:

... one would have to adopt a very radical materialism indeed, rejecting not only the dualism of substances, but also the dualism of properties, and even the distinction of first- and third-person aspects or ways of identifying the sensible qualities, as well as the notion of a way in which conscious states seem to us when we are in them, as opposed to their place in the physical scheme of things. Thus one would have to eliminate phenomenal qualia, or reduce them in a most extreme way to physical qualities.⁵²

However, the materialism suggested by our open possibility can retain the distinction between first- and third-person ways of identifying the sensible qualities, and also the notion of a way in which conscious states seem to us when we are in them. For despite the discrepancy between the real qualitative nature of phenomenal properties and how they are introspectively represented, there is a first-person, introspective point of view on phenomenal properties, and a way they appear to us when we are in the states that have them. True, the real qualitative nature of those properties would be accessible from the third-person point of view, so the first-person perspective does not provide genuine information about the qualitative nature of those properties that is not accessible from the third-person perspective. But this is not enough to make the materialism in question a radical one, since a claim of this sort would be required for any materialism.

demands, and that for this reason physicalism is false. But now we can see that on the open possibility under investigation, Mary would be able to produce these explanations.

⁵² Robert Adams, "Flavors, Colors, and God," p. 259.

At the same time, denying that a qualitative nature that resembles introspective phenomenal modes of presentation is correctly attributed to phenomenal properties might well not amount to eliminativism for phenomenal properties. One could, in agreement with Galileo, argue for eliminativism about temperature as a property of physical objects, on the grounds that the temperature of physical objects does not resemble our sensory representation of it.⁵³ An Edenic content of our temperature concept could even be defined that applies only to a property that resembles temperature sensations -- and it might then be pointed out that there is no actually instantiated property to which this concept correctly applies. But, as history has shown, highly plausible non-eliminativist options for temperature itself remain. On the open possibility that I have been discussing, non-eliminativist options also remain for phenomenal properties. *Something* that many believe to exist would be eliminated – certain features that are accurately represented introspectively. Indeed, one might define a notion of the Edenic content of phenomenal concepts that would correctly apply only to such features, which would then correctly apply to no properties that are actually instantiated. But this is not to say that phenomenal properties would thereby be eliminated, or that our phenomenal concepts fail to apply to anything real, for they might have an ordinary content that does. Even then, the Edenic content might still function as a regulative ideal, on the model for color concepts Chalmers develops.

6. The big picture.

⁵³ Galileo Galilei, *The Assayer*, in *Discoveries and Opinions of Galileo*, pp. 274-7.

Against the knowledge argument I have contended that because it is an open possibility that the correct account of introspective representation is causal, it is also an open possibility that introspective representations of phenomenal properties are qualitatively inaccurate. Introspection represents phenomenal properties as having a certain qualitative nature, but the attribution of this qualitative nature to phenomenal properties might be incorrect. As a result, it may be that the real nature of phenomenal properties is straightforwardly physical, and derivable from what Mary knows before emerging from the room. Arguably, Kant would have endorsed the open possibility of this kind of qualitative inaccuracy, since for him any representation of the qualitative nature of an ultimately real entity might not – or in fact does not – represent that entity as it really is.⁵⁴ By contrast, rationalists like René Descartes maintain that for both the metaphysically real mental and material realms we have representations that are qualitatively accurate. Descartes held that for the mental they are the clear and distinct introspective representations of our own psychological states.⁵⁵ Perhaps we should say that both the Cartesian and Kantian options, broadly construed, are live open possibilities; neither has been defeated. But as long as the Kantian view remains standing, the knowledge argument against physicalism faces a serious challenge.

⁵⁴ Immanuel Kant, *Critique of Pure Reason*, e.g., Bxxiv-xxvii.

⁵⁵ René Descartes, *Meditations on First Philosophy*, in *The Philosophical Writings of Descartes*, v. 2., esp. pp. 16-23; *Principles of Philosophy*, in *The Philosophical Writings of Descartes*, v. 1, e.g. p. 247.

