

**Review of Torin Alter and Sven Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism***

James Trafford  
University of East London  
jamestrafford1@googlemail.com

**Alter, Torin and Walter, Sven, (eds.), (2007) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Philosophy of Mind Series, Oxford: Oxford University Press, 349 pp., ISBN 9780195171655**

*Phenomenal Concepts and Phenomenal Knowledge* is an edited volume of new essays relating to the debates around phenomenal experience in philosophy of mind. Alter and Walter provide an excellent introduction to the volume, producing a well edited collection of papers that represent some of the most interesting and cutting edge work in the field, and together provide a subtle and complex overview of the contemporary theoretical landscape. In addition, as many of the papers refer to others within the volume, they provide an excellent opportunity for in depth and complex debate between some of the leading theorists at work today. The first part of the collection concentrates on phenomenal knowledge, with essays based around Frank Jackson's *knowledge argument*. The latter half contains essays on the broader notion of phenomenal concepts, concentrating mostly on anti-physicalist arguments and the popular a posteriori physicalist response.

First, Mary sets the scene: On the original knowledge argument (Jackson, 1982), the ineluctable Mary is a color-scientist, trapped in a black and white room from birth, where she learns everything that can be known about color at the microphysical, macrophysical and functional levels. Thus, before Mary leaves the room, she knows all the physical facts related to color experiences without ever experiencing color. When Mary leaves the room, it seems intuitively correct that she will learn something new when she experiences color. Since Mary possessed all the physical facts, and since this new knowledge cannot be deduced from those facts, then it seems that physicalism does not hold for facts relating to conscious experiences. We can call this the *new knowledge* intuition.

Underlying the new knowledge intuition is the thesis of a priori physicalism: if physicalism is true, then, given all the appropriate physical facts, all facts will be derivable a priori. This, together with a minimality thesis, which proposes that any minimal physical duplicate of our world is a duplicate *simpliciter* of our world, suggests that the following (simplified) entailment will hold: Every term *M* expressing a true fact about the world can be derived from the relevant physical facts *P*, so that  $P \rightarrow M$  is both conceptually true and metaphysically necessary. More importantly, if physicalism is true, then for every phenomenal truth *Q*,  $P \rightarrow Q$  will similarly be both conceptually true and metaphysically necessary. By definition, a

priori physicalism is true if knowing all the physical truths a priori entails knowing all the truths about the world simpliciter.

The knowledge argument puts pressure on this second entailment by promoting the existence of an epistemic gap between physical facts and phenomenal facts. If the epistemic gap warrants a metaphysical gap, then physicalism is falsified. That phenomenal facts cannot be a priori entailed from the physical facts is agreed upon by both a posteriori physicalists and anti-physicalists because true phenomenal facts cannot be construed to be causally or functionally fixed in the way ordinary facts can be. Hence, a posteriori physicalists have to deny that the epistemic gap supports a corresponding metaphysical gap by appealing to the special nature of the phenomenal concepts involved.

As Alter and Walter point out in their introduction, two related questions emerge, which are addressed throughout the volume:

1. Could a proper understanding of phenomenal concepts / knowledge show that there is or is not an epistemic gap?
2. Could a proper understanding of phenomenal concepts / knowledge show that there is or is not a metaphysical gap? (p. 5)

The second section of the volume deals with the latter question, primarily with the *phenomenal concept strategy*, which attempts to provide an explanation of phenomenal concepts that can explain the epistemic gap without entailing a metaphysical gap, and simultaneously answering the relevant intuitions regarding phenomenal experience. There are several accounts of phenomenal concepts in the literature: as sui generis (Chalmers, 2003); as demonstratives (Horgan, 1984; Perry, 2001); as recognitional (Loar, 1990); and as quotational (Papineau, 2002). Levin, in “What is a Phenomenal Concept?” closely follows Loar’s account: Phenomenal concepts are conceptually independent representations that a subject can acquire only by having the experiences that they denote; like demonstratives, they pick out their referents directly. Levin points out that a number of physicalists have recently criticized this account (including Block and Papineau in this volume). The criticisms suggest that the demonstrative account is problematic. The problem is: phenomenal concepts cannot both directly refer and be robust enough to account for the substantive knowledge that Mary acquires without invoking a non-physical mode of presentation. Levin provides answers to these problems, arguing that physicalists ought to refuse the idea that phenomenal concepts should explain our “acquaintance” with phenomenal qualities since a type-demonstrative account without appeal to any mode of presentation is sufficiently robust.

Papineau, in “Phenomenal and Perceptual Concepts,” develops a contrasting view that takes acquaintance into account, suggesting that there is an intimate relation between phenomenal concepts and phenomenal qualities: “...phenomenal concepts have the very peculiar feature of *using* the experiences they refer to” (p. 131). Papineau develops his previous quotational account (2002), which suggested that phenomenal concepts have the structure “*the experience: –*”, where “–” is filled with an actual perceptual experience or an imaginative re-creation thereof (p. 120). Phenomenal concepts can thus be understood as literally “quoting” their referents. Here, Papineau refines the account, suggesting that “...phenomenal concepts are simply special cases of perceptual contents” (p. 122). This idea expands on Jesse Prinz’s work on sensory templates: “The reason that Mary’s new concept depends on

experience is that it requires a sensory template, and her acquisition of this template depends on her visual system having been activated previously by some red surface” (p. 127).

Block, in “Max Black’s Objection to Mind-Body Identity” makes a similar argument that phenomenal concepts are their own mode of presentation. He employs this view to set the property dualism argument against identity theory proposed by both Max Black and Stephen L. White in “Property Dualism, Phenomenal Concepts, and the Semantic Premise” also in this volume. Simply put, the argument is concerned with the way necessary a posteriori identity claims, flanked by natural kind terms and physical terms (such as, water = H<sub>2</sub>O), have an appearance of contingency. This can be put down to the distinctness of the concepts flanking the identity statement, and the way that ordinary natural kind terms, such as water, have associated properties that function as a contingent mode of presentation. The a posteriori physicalist wants to maintain that the identity claim, pain = C-fibers firing (CFF), is an example of the necessary a posteriori, but the apparent contingency cannot be explained in the same manner as ordinary kind terms. This is because everyone agrees that whatever seems to be pain, is, essentially, pain. Block suggests that the apparent contingency of the identity of pain with CFF can be explained by appeal to the fact that there are two distinct modes of presentation at play: On the one hand, we have pain = CFF, given as a specific brain state, and on the other, as the ordinary presentation of pain. White argues that invoked modes of presentation cannot play all the roles they need to in order for the physicalist argument to stand.

White’s argument can be boiled down to a single claim: Since modes of presentation correspond to features or properties of items in the world (by virtue of which a mode of presentation picks them out), then there is a lack of referential space. If both pain and CFF are essential properties of the referent, then the identity statements ought to invoke the same mode of presentation. This is largely due to the Fregean constraint on coherence:

What is required to explain fully the a posteriori character of the mental-physical identities is not just that the concepts flanking the identity sign have different conceptual roles. What is required is an explanation of how the subject who claims sincerely not to believe such an identity takes the world to be. (p. 212)

According to White, invoking direct reference or acquaintance will not satisfy this requirement because direct reference is too thin to yield robust data for a posteriori identities. White argues that on the phenomenal concept strategy, property identities must connote coextensive properties of the referent. If, however, both pain and CFF connote essential properties of the referent, then there is no room for an apparent identity contingency that is not also a real contingency. (p. 225). Block attempts to defuse these problems by suggesting that the argument conflates two distinct modes of presentation: the conceptual (CMOP), having to do with semantics or inferential roles; and the metaphysical (MMOP), having to do with the property of the referent. This approach seems promising, and deserves further development, particularly in relation to its wider application to Frege problems and recent advances in two-dimensional semantics.

Nonetheless, as White points out, the key problem for acquaintance or direct reference theories of phenomenal concepts seems to come down to a lack of

conceptual space. On these accounts, a phenomenal concept is constituted by the phenomenal experience to which it refers. For example, Papineau states that "...phenomenal concepts are too close to their referents for it to seem possible that those same concepts could refer to something else" (p. 132). The closer phenomenal concepts are supposed to be to their referents, the more unlikely they are to mislead about the nature of those referents. Levine makes a similar point in "Phenomenal Concepts and the Materialist Constraint," where he argues that neither the direct reference account nor the quotational account can explain significant aspects of our intuitions regarding acquaintance:

One might say that there now is a second explanatory gap: between implementations of cognitive architecture and whatever it is about phenomenal concepts – in my terms, that they afford genuine cognitive presence to phenomenal properties – that is responsible for the original explanatory gap. (p. 165)

Moreover, even if it is granted that the physical and the phenomenal could seem different, there is little in the account of phenomenal concepts themselves that is persuasive regarding why they, in fact, turn out to be the same. Recall that on Papineau's account, a phenomenal concept does not simply directly refer, it literally uses its referent: There is no appearance / reality distinction for phenomenal concepts themselves. If this is the case, then we would seem to have an unmediated grasp of phenomenal properties, and there is a lack of conceptual space between those properties and physical concepts.

It is this kind of problem that Chalmers exploits when he sets up his "master argument" against the phenomenal concept strategy (p. 173). Chalmers poses a dilemma by arguing that phenomenal concepts cannot jointly satisfy both physicalistic explanation and our epistemic situation in relation to phenomenal qualities. Let *C* be the thesis that humans possess phenomenal concepts in all the relevant ways discussed above (explaining our epistemic situation to consciousness; explaining epistemic gaps; that there is a physical explanation of phenomenal concepts) (p. 172). Chalmers develops the zombie argument (conceivable beings who are physically identical to humans but lacking consciousness) by asking whether or not, given all the physical facts, *P*, is  $\sim C$  conceivable. The master argument goes as follows:

1. If  $P \& \sim C$  is conceivable, then *C* is not physically explicable.
  2. If  $P \& \sim C$  is not conceivable, then *C* cannot explain our epistemic situation.
  3. Either *C* is not physically explicable, or *C* cannot explain our epistemic situation.
- (p. 174)

Papineau responds to the dilemma. Taking the first horn (premise 1), he simply explicates a version of the phenomenal concept strategy by appealing to second-order phenomenal concepts (p. 138). However, it is Papineau's response to the second horn (premise 2) that is interesting, because he does not invoke the phenomenal concept strategy. In fact, he states that:

...type-B [a posteriori] physicalists should bite the bullet and say that the thing that differentiates us from the [...] zombies doesn't make any difference to the explanatory significance of phenomenal concepts. (p. 142)

In this sense, Chalmers' master argument would be guilty of begging the question by implicitly invoking a special epistemic relation to phenomenal qualities that a posteriori physicalists deny. He does this by arguing that any being sharing our physical properties will automatically share our conscious properties (p. 141). Since Papineau invokes the physical-functional description at this higher-level in order to explain our epistemic situation, one might wonder why this doesn't also operate at the lower level of description. Given the associated problems with phenomenal concepts, and the readiness to bite this bullet, why not be far more austere when it comes to conceivability arguments?

Also in this second section, John Hawthorne, in his essay "Direct Reference and Dancing Qualia," argues against Chalmers' own understanding of phenomenal concepts by showing that for phenomenal concepts, as for proper names, "...direct reference semantics is incompatible with the thesis that the a priori of a thought token turns on the proposition it expresses" (p. 208). Martina Nida-Rümelin, in "Grasping Phenomenal Properties," argues that phenomenal properties are "grasped" via phenomenal concepts in the Cartesian sense: We have direct knowledge of their essential properties. This claim is similar to acquaintance theories, but by arguing for the cognitive independence of physical and phenomenal concepts, she provides an interesting anti-physicalist argument.

Turning to the first section of the volume, Frank Jackson, who formulated the knowledge argument, now represents a far stricter position on conceivability arguments. Jackson's current view is a combination of representationalism with the *ability hypothesis*, the latter defended in this volume by one of its originators, Laurence Nemirow, in "So This Is What It's Like: A Defense of the Ability Hypothesis." On the ability hypothesis, the new knowledge intuition is explained by the suggestion that Mary does not gain factual knowledge on leaving the room, rather, she gains abilities to recognize, imagine, and remember her experiences. The major problem with Nemirow's defense, as with the hypothesis in general, is its counter-intuitive nature. Since Mary's knowledge is usually associated with occurrent token phenomenal states, rather than the acquisition of types that have the associated abilities, the hypothesis fails to explain the substance of the new knowledge intuition.

Daniel Dennett, in "What RoboMary Knows," points out that he has consistently argued that a materialist theory of consciousness may well be very counterintuitive (p. 30). This is refreshing against the resurgence of interest in a priori analysis and the concessions made to the intuition of acquaintance by many a posteriori physicalists. Nonetheless, it remains the case that philosophers should attempt to explain our intuitions while providing independent reasons for motivating counter-intuitive alternatives. Yet, like Nemirow, it is here that Dennett seems to falter. Underlying Dennett's RoboMary thought experiment is the identification of phenomenal knowledge with the acquisition of abilities, following the ability hypothesis. Surpassing the hypothesis, however, Dennett attempts to motivate the claim that these abilities can also be acquired before Mary's experiences outside of the room. This seems unnecessarily strong since advocates of the ability hypothesis have supposedly already shown how they avoid anti-physicalist threats. Nonetheless, Dennett's argument that RoboMary will be able to come to phenomenal experiences

as a result of her complete physical knowledge, admirably bites the bullet with regard to the remarkable effect that knowledge of a completed physical science might have. In doing so, however, Dennett is potentially left without an account of the relationship between the phenomenal and physical knowledge tout court. That is to say, even if current philosophy is simply conservative conceptual anthropology, hostage to the advances of the sciences, aren't we still owed an account of how our natural language philosophizing and our phenomenal experiences are to fit within this scientific vision? In other words, even if ultimately our commonsense intuitions and vocabulary regarding the phenomenal will be significantly revised, we would still need to develop a revisionary account that integrates those that *manifest image* with the *scientific image*, to borrow Wilfrid Sellars' terminology.

Jackson's attempt to provide an account of the relationship between the physical and the phenomenal rests upon a representationalist component. He expounds this in detail in "The Knowledge Argument, Diaphanousness, Representationalism," in tandem with Alter's critical paper, "Does Representationalism Undermine the Knowledge Argument?" Jackson accepts a priori entailment from physical facts to all truths about the world, including phenomenal truths. Jackson's goal is to tackle the intuition that the new knowledge intuition directly entails that Mary learns the truths about a property when she first experiences the color red. As stated, the phenomenal concept strategy attempts to motivate the claim that it is new concepts rather than knowledge about new properties that Mary acquires. Jackson argues that this is an unsatisfactory response because Mary seems to acquire a new way of grouping experiences together (p. 53):

...Mary's new concept seems to correspond to a new way for experiences to be alike, one that nowhere appears in the physicalist's picture; and if this is right, there are properties that fail to appear in that picture. (p. 53)

Jackson calls this the "new similarity" contention, and proposes representationalism as its solution. The new kind of similarity holding between experiences that Mary learns about can be understood by appealing to strong representationalism, which maintains that:

...how an experience represents things as being exhausts its experiential nature [...] Change an experience qua kind of experience it is and you ipso facto change how it represents things to be. (p. 57)

On this view, the new similarity Mary learns is not put down to any instantiated property but to an intensional property: To say that Mary experiences 'red' is simply to talk of her experiencing representing 'red.' Following other representationalists, Jackson suggests that experience is phenomenologically transparent; it is diaphanous in the sense that we do not experience, "experience" as such. Mary's experience does not have to do with an instantiated property of 'red,' because "...the experience's properties are one and all the properties of how things are represented to be [...] the experiences properties qua kind of experience it is" (pp. 60-61). Jackson considers this to be a viable answer to the knowledge argument, since Mary does not acquire an enlarged range of properties that she holds to be instantiated

in our world (p. 63). Rather, Mary's new kind of experience is exhausted by representation: "...properties of how things are being represented to be are not instantiated properties; talk of properties of intentional objects is a mere manner of speech" (p. 63).

Jackson's approach is promising, but there are significant areas that need further development. Alter's paper addresses the issue that since Mary's experience must have content because it represents things to be, then this content will need an *a priori* (given Jackson's position) physicalist explanation. That is to say, Jackson must hold that the facts about the content of Mary's experience must be *a priori* derivable before she leaves her black and white room. Since Jackson has effectively removed the problem of having to derive an instantiated property, it is plausible for Jackson to turn to the naturalistic construal of content. Alter provides an account of Jackson's (2003) method for this process in three stages: i. Mental representation is physically explicable; ii. There are five general features explaining phenomenal representation; iii. All five features are physically explicable. Alter suggests that though this is plausible, it does not satisfy the strong intuition that Mary will learn something new about phenomenally representing red on leaving the room (p. 69). For example, stages i. and iii. can be granted without granting ii. because:

One could say that seeing the tomato allows Mary to eliminate epistemic possibilities concerning how seeing red represents: possibilities that she cannot eliminate, or fully understand, before she leaves the room, despite her comprehensive physical knowledge. (p. 70)

Therefore, Mary's epistemic progress in leaving the room would not satisfy our intuitions without appealing to instantiated properties. Furthermore, Jackson has to invoke something like the ability hypothesis in order to explain the new similarity contention, since representationalism does not explain knowledge of new instantiated properties, or acquisition of concepts: It is ability to group together experiences. Here, Alter allows that this would be an independent reason for rejecting the knowledge argument, but then it would be the ability hypothesis, rather than strong representationalism that does the substantive work in overcoming the argument. Jackson would then seem to run into familiar problems with the ability hypothesis in invoking types, rather than tokens, of phenomenal states. This is to say, the abilities that Mary is supposed to acquire are put to use grouping experiences together after the occurrent token of experience, and therefore do nothing to explain our knowledge of the phenomenal experience as she experiences it. This is a strong intuition that the new similarity contention does not explain, and does not seem to be forthcoming from the kind of account Jackson offers. The only option available to Jackson seems to be to bite the bullet of strong representationalism and to argue that Mary can know what it is like to experience 'red' in an exhaustive sense before leaving her room; in so doing, he would deny the intuition that the knowledge argument made so compelling and would come to a position that is as counter-intuitive as Dennett's. This is not to say that a counter-intuitive position is inherently false, but, again, since Jackson is intent on preserving something like *a priori* conceptual analysis as a strong philosophical method, it seems like we are owed an account of such a strong intuition.

Also in the first section, Knut Nordby, in “What is This Thing You Call Color?” discusses the intuitions underlying the knowledge argument in terms of vision science and color blindness.

The discussions in this volume are at the centre of debates around the relation between consciousness and the physical world, and they succeed in both developing the conceptual landscape and throwing up new issues that will be pondered over in the coming years. This is a compelling volume that is thoughtfully constructed and is essential reading for anyone with an interest in the contemporary debate around consciousness in philosophy of mind.

## **References**

- Chalmers, D. (2003). The Content and Epistemology of Phenomenal Belief. In Q. Smith and A. Jokic (Eds.), *Consciousness: New Philosophical Perspectives* (pp. 220-72). Oxford: Oxford University Press.
- Horgan, T. (1984). Jackson on Physical Information. *Philosophical Quarterly*, 34, 147-83.
- Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly*, 32, 127-36.
- Jackson, F. (2003). Mind and Illusion. In A. O’Hear (Ed.), *Minds and Persons: Royal Institute of Philosophy Supplement 53* (pp. 251-71). Cambridge, UK: Cambridge University Press.
- Loar, B. (1990). Phenomenal States. In J. Tomberlin (Ed.), *Philosophical Perspectives 4: Action Theory and Philosophy of Mind* (pp. 81-108). Atascadero, CA: Ridgeview.
- Papineau, D. (2002). *Thinking About Consciousness*. Oxford: Oxford University Press.
- Perry, J. (2001). *Knowledge, Possibility, and Consciousness*. Cambridge, MA: MIT Press.