Mary Meets Molyneux: The Explanatory Gap and the Individuation of Phenomenal Concepts

CYNTHIA MACDONALD
University of Canterbury

It is widely accepted that physicalism faces its most serious challenge when it comes to making room for the phenomenal character of psychological experience, its so-called what-it-is-like aspect. The challenge has surfaced repeatedly over the past two decades in a variety of forms. In a particularly striking one, Frank Jackson considers a situation in which Mary, a brilliant scientist who knows all the physical facts there are to know about psychological experience, has spent the whole of her life in a black and white room. He asks,

What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. (Jackson 1986: 130)

Jackson’s version of the challenge doesn’t mention any problem about explanation. However, Levine (2001) takes it to show that there is an explanatory problem for physicalism.

...Jackson’s story does show something important about Mary’s epistemic situation; in particular, her ability to explain qualia in physical terms. For if Mary could really explain the character of sensory experience by reference to the underlying physical processes, then it seems that she shouldn’t learn anything new when she finally experiences red for herself. She should have expected it to be like that. The fact that it seems so clear that she would learn what it’s like to experience red is testimony to the explanatory gap that separates physical theory and conscious experience. (Levine 2001: 77)
Philosophers have responded differently to the explanatory problem that the challenge seems to generate. Some (e.g., Chalmers 1996) have claimed that the existence of an explanatory gap between the physical domain and the domain of conscious experience shows that dualism is true. Others (e.g., Loar 1990, Levine 1983, 1998, 2001, Papineau 2002), while acknowledging the existence of such a gap, have argued that it poses no threat to physicalism.

My interest in this paper is to address a cluster of responses of the second kind (ones that acknowledge the existence of the gap and defend physicalism against it), which employ a common strategy. The strategy appeals to the view that (1) psychophysical identity statements are necessary and a posteriori, and (2) the deployment of phenomenal concepts in some such statements leads to an explanatory gap because phenomenal concepts form a special kind of direct recognitional concept.

Call this the Phenomenal Concept, or PC strategy. In what follows, I argue that one well-known way (Tye 1999, 2000; Harman 1990) of developing the PC strategy is seriously defective. Philosophers have different views about what phenomenal concepts are. Some (Tye 1999, 2000; Harman 1990) take them to be concepts of properties, whose possession requires the undergoing of phenomenal experiences. Others (Loar 1990) take them to be concepts of properties of phenomenal experiences. My target here is the PC strategy developed in the first sort of way, according to which phenomenal concepts are ones whose possession requires the undergoing of phenomenal experiences.

In sections one and two I outline the explanatory gap problem and show how the PC strategy is supposed to deal with it. In section three I argue that considerations about concept individuation tell against the strategy. I go on to suggest an alternative strategy for handling the explanatory gap problem that is similar to the defective one in the way in which it attempts to meet the challenge to physicalism, but which does not suffer from its principal defect. The alternative endorses the claim that the explanatory gap problem is a conceptual one, but rejects the claim that the gap arises because psychophysical identity statements employ phenomenal concepts. I then defend the alternative strategy against three obvious objections. If it is right, there is another way of defending a posteriori physicalism against explanatory gap objections by appeal to considerations about concepts that does not rely on claims about the distinctive nature of phenomenal concepts.

1. What Is the Explanatory Gap Problem?

Let me begin by assuming the truth of physicalism, which I take to be the view that all mental phenomena as well as mental kinds (such as the kind, pain) are identical with physical phenomena and kinds. It licenses, as true, identity statements linking phenomenal with physical kinds, such as ‘Pain is
identical with C-fibres’ firing’, or ‘Brain State B is identical with the experience of red’. One way in which to interpret the question ‘How can soggy grey matter feel like something?’ is as asking how it could be, say, that the kind, experience of red—something that feels like something—is identical with the kind, Brain State B.7

It is standard for physicalists to respond by saying that there is no answer to this question, but no need for one either (Papineau 1993). In general, identity statements are not statements of the sort that need explaining. What might need explaining is why one would be justified in believing a true identity statement—what evidence one might have for its truth.

One can imagine, for example, a person who is ignorant of chemistry and the chemical facts about water asking how it could be that ‘Water is identical with H2O’ is true. We can explain this to her by appealing to the facts about chemical elements, compounds, and how these combine to form substances with observable properties such as odourlessness, transparency, and so on. But once we have marshalled forth all the evidence available for thinking that water is H2O, there is nothing more that needs to be explained. If a person were to persist in asking how it could be that water is H2O we would think that they had simply failed to understand the explanation, or the identity claim, or both.

Similar remarks apply to identity statements like ‘Pain is identical with C-fibres’ firing’ or ‘Brain State B is identical with the experience of red’. It is true that we are not currently in a position to identify those physical state types or kinds with which phenomenal kinds are identical. This might tempt us to think that there is a further mystery in the case of phenomenal kinds that doesn’t attach to physical kinds in general. But there isn’t. What is missing here is simply evidence for thinking that the relevant identity statements are true.

Implicit in this response’s appeal to the claim that true identity statements need no explanation is a second and more important claim. It is that the intuition that there is an explanatory gap in the psychophysical case is due to lack of information about the physical state kinds that are the likely candidates for identification with phenomenal state kinds. This requires the view that identity statements linking mental and physical kinds are a posteriori rather than a priori true, this feature giving them the appearance of opening up an explanatory gap that needs to be filled.

The initial response to the explanatory gap problem, then, is that true identity statements need no explanation, but we might need to explain why we are justified in believing them; which in turn appeals to:

(U) True identity statements are ungappy.

and

(AP) Psychophysical identity statements are necessary and a posteriori true (a posteriori physicalism).
Does the strategy work? Although the matter is controversial, I think that the general consensus has been ‘no’. Both claims involved in it have come under scrutiny. Against (U) it has been argued that it is false that no identity statement needs explaining (Levine 1998). Identity statements can be either ‘ungappy’ or ‘gappy’. ‘Ungappy’ ones don’t need explaining, but ‘gappy’ ones do; and psychophysical identity statements are ‘gappy’. With regard to (AP) it has been argued that even if it is true that identity statements linking mental and physical kinds are \textit{a posteriori} necessary truths, this cannot suffice to solve the problem of the explanatory gap (Stoljar 2000).

In what follows I want to focus on physicalists’ appeal to (AP) to solve the problem of the explanatory gap, since, irrespective of whether or not (U) is true, there seems to be agreement amongst physicalists that identity statements linking phenomenal and physical kinds \textit{appear} to be ‘gappy’. (AP) plays a major role in the attempt to explain this appearance. So the success or failure of physicalism depends crucially on whether the attempt succeeds. According to the strategy under consideration, this depends on the truth of the further claim

\begin{itemize}
  \item[(PC)] Phenomenal concepts form a kind of direct recognitional concept.
\end{itemize}

2. \textbf{A Posteriori Necessary Identities, Phenomenal Experiences, and Phenomenal Concepts}

Many \textit{a posteriori} physicalists account for the appearance of an explanatory gap by claiming that it arises from the distinctive nature of our phenomenal concepts alone. In general, explanation invokes two items: kinds, and concepts of those kinds. The response is that if what is at issue is the relation between phenomenal and physical properties or kinds, then the appearance of an explanatory gap is mere appearance. If, on the other hand, what is at issue is the relation between phenomenal and physical \textit{concepts}, then it is not a problem. On the contrary: \textit{a posteriori} physicalism \textit{requires} that there be a gap between mental and physical concepts.

It is recognized, though, that this, in itself, is insufficient to account for the distinctive nature of the ‘gap’ which \textit{a posteriori} physicalists and their opponents both see in psychophysical identity statements. The problem is that these statements are unlike other true \textit{a posteriori} identity statements, such as ‘Water is identical with H\textsubscript{2}O’. Once we have advanced all the evidence available for thinking these other statements are true, there is no residual worry about how it could be that they are true. But this is not so in cases of true psychophysical identity statements, such as ‘Brain State B is identical with the experience of red’. Even after all the available evidence is in for thinking that these statements are true, the worry persists as to how it could be that they are true.
According to the PC strategy, the reason why the ‘gap’ persists in the case of psychophysical identity statements of the latter sort is that they employ a distinctive type of concept, the direct recognitional type. A direct recognitional concept is one whose possession requires undergoing experiences of instances that fall under it and whose possession enables its subject to recognize instances that fall under it ‘directly’, as instances that fall under it, without need of mediation by other descriptive information about such instances. If, as (PC) states, phenomenal concepts are a distinctive type of direct recognitional concept, particular experiences of phenomenal kinds is a necessary condition on the possession of those concepts.8

A standard way of developing and defending (PC) is to argue that there is a distinctive mode of awareness, a subjective one, associated with phenomenal concepts that other non-phenomenal concepts fail to have.9 Thus consider, for example, the following claims:

To possess the phenomenal concept red, for example, is to possess a simple concept that has been acquired by undergoing experiences of red (barring neurosurgery to induce the state or a miracle) and that not only disposes one to form a visual image of red in response to a range of cognitive tasks pertaining to red but also is brought to bear in discriminating the experience of red from other colour experiences in a direct and immediate manner via introspection. (Tye 1999:712)

Each phenomenal concept is thus tied to a particular experience-specific perspective occupied by the possessor of the concept. As the experiences vary, so too do the phenomenal concepts. (Tye 1999: 709)

Tye here speaks of the way in which phenomenal concepts are acquired, viz., by undergoing phenomenal experiences; but it is clear that he considers the having of phenomenal experiences to be a necessary condition on an individual’s possession of phenomenal concepts (“barring neurosurgery or a miracle”).10 For evidently, this is what is required for such concepts to have their distinctively first-personal nature.

According to the PC strategy under consideration, concepts like red and experience of red are phenomenal concepts. This being so, the reason why Mary, when confined to her black and white room, doesn’t know what it is like to experience the colour of a ripe tomato is that she lacks the experiences required for possession of the phenomenal concepts red and experience of red and for any beliefs in which those concepts are embedded. What Mary acquires on release from her room is not knowledge of a new fact about the non-conceptual world. What she acquires are new concepts.11

How is the strategy meant to solve the explanatory gap problem? Because the phenomenal concepts red and experience of red enable their subjects to be put in direct contact with their own experiences of red in a way in which
no concept of Brain State B could do, it puts them in a position to recognize directly and in an immediate way their own phenomenal experiences of red. Possession of the concept Brain State B could not put any person in a position to recognize directly and in an immediate way their own visual experiences of red. So, even having met the particular experience requirement on the possession of red and experience of red, one might wonder how it could be that the experience of red is identical with Brain State B.

Understood in this way, the epistemic gap between phenomenal and physical concepts misleads one into thinking that because one’s mode of access to one’s experiences is radically different from one’s mode of access to physical states, what one has access to when one deploys a phenomenal concept is radically different from what one has access to when one deploys a physical concept. But according to the present strategy, this is an illusion, created by the distinctive first-personal nature of phenomenal concepts.

Does the strategy solve the problem of the explanatory gap? Well, in one sense, it plainly does not. A conceptual gap is a gap. However, physicalism never purported to close the gap between phenomenal and physical concepts. It took its job to be an ontological one alone. So the sense in which the strategy fails to close the explanatory gap shouldn’t worry physicalists about the phenomenal, any more than the gap between the concept water and the concept H₂O should worry those concerned to explain the relationship between water and H₂O. Such conceptual gaps are rife in areas other than ones in which phenomenal and other mental concepts figure, in domains that are squarely within the realm of the physical, as the water/H₂O case illustrates. And the residual feeling of mystery associated with psychophysical identity statements can be explained by appeal to the very distinctive first-personal nature of phenomenal concepts, as compared with concepts of natural kinds.

3. Mary Meets Molyneux

Nevertheless, I think that the strategy fails. The problem with it is deep-rooted and very general, and has to do with its commitment to both (PC) and the view that concepts like red are phenomenal concepts. It concerns the question of what it is to be a direct recognitional concept, and with it the issue of concept individuation. To see why the strategy fails, we need to appreciate certain conclusions about concept-individuation that are to be drawn from discussions of Molyneux’s question.

Ever since Locke introduced Molyneux’s question in his Essay Concerning Human Understanding, philosophers and psychologists have been fascinated by issues concerning relationships that the different sense modalities bear to one another and to concepts of objects and properties in the mind-independent world.¹² Consider an individual who was blind from birth but whose other sense modalities, specifically the haptic one, were fully functional, and thus
capable of delivering, say, tactile experiences of shapes such as squares and circles. Suppose that he were suddenly to acquire his sight as an adult. Molyneux’s question was, would that individual be capable of identifying visually the things and properties in the world around him that he had previously been able to identify through the use of his other sense modalities such as the haptic one? Further—and this is perhaps more our present question than Molyneux’s—what, if anything, would a positive or negative answer to this question show about such concepts as those of, say, shapes, which are typically acquired on the basis of both tactile and visual experiences?

Molyneux’s question can be interpreted in many ways, but one of these is particularly relevant to the present discussion. According to it, the question is not one about whether the subject—let’s call him Harry—would or would not need to learn how to apply shape concepts visually. We can take it as given that he would need to do this. Rather, the question is, what has he learned how to do? Has he learned how to apply a host of newly acquired, visual concepts, to things in the visual world, which he might then connect with the things identified by means of touch and which fall under tactile concepts, such as, say, tactile squareness? Or has he learned how to apply, to things experienced visually, a single, cross-modal concept, the concept square, whose possession by him was originally associated with tactile experiences? If the former, the answer to Molyneux’s question is ‘no’: such terms as ‘square’ are ambiguous depending on whether they are used in visually-based or haptically-based judgements (i.e., there is no one concept common to both sorts of judgements), and Harry’s ability to identify square shapes by means of touch will not enable him to extend his concept, square, to squares identified visually. This is usually taken to show that such concepts as square are in fact modality-specific. If the latter, the answer to Molyneux’s question is ‘yes’, and Harry’s visual identification of shapes in his environment need involve no more than the ability to extend his haptically-based shape concepts such as square to squares experienced visually. This is usually taken to show that such concepts are cross-modal.

This question and its answer have an important bearing on the present discussion. For the PC strategy is effectively pursuing the same kind of line with regard to phenomenal concepts as those who wish to answer ‘no’ to Molyneux’s question with regard to a concept such as square. (In fact, as we shall see, it is pursuing an even more radical line.) A negative answer to Molyneux’s question encourages us to think that that a person blind from birth, such as Harry, fails to possess the modality-specific visual concept, square, even though he might possess the modality-specific haptic concept, square, in virtue of having tactile experiences of square shapes in his environment. Similarly, the PC strategy encourages us to think that that Mary fails to possess the modality-specific visual concept red, despite the fact that she is not blind, and although she might possess the (non-phenomenal) concept red by some other means (say, by analysis of wavelengths of light). According to the
PC strategy, Mary lacks the phenomenal concept *red*, because it is a necessary condition of her possession of that concept that she undergoes visual experiences of red.

In pursuing the line taken by those who give a negative answer to Molyneux's question, the PC strategy is relying on the assumption that it is sufficient to establish that concepts are distinct that they are tied to different sense modalities. But a negative answer to the question, and with it endorsement of the assumption on which the answer is based, is objectionable on both empirical and philosophical grounds.

Those who object on empirical grounds need not deny that concepts such as *square* are typically acquired on the basis of both visual and tactile experiences, nor that the various sense modalities deliver up distinctive experiences not available through the exercise of other sense modalities. What they do deny, however, is that there is no common concept employed in both visually and haptically based judgements, a concept that applies to square things in the world. Empirical work on very young infants, for example, strongly suggests that they engage in a large degree of cross-modal matching without the benefit of associations acquired through learning between the sense modalities (Meltzoff 1993, Steri and Pecheux 1986, and Steri 1987). Specifically, they have an impressive ability to integrate and match felt and seen shapes, sounds from seen and heard lip movements, imitating seen gestures, and so on. This work supports the view that concepts such as *square* are cross-modal, rather than modality-specific. Correspondingly, it suggests that the assumption that it is sufficient for concepts to be distinct that they are tied to different sense modalities is false.

This conclusion is further supported by philosophical considerations. Many supposedly direct recognitional concepts—for example, the concept *square*, the concept *rough* (as in *rough surface*), the concept *loud* (as in *loud noise*)—are ones whose possession, it is plausible to suppose, requires the undergoing of particular experiences. But this leaves it an open question whether the experiences required for their possession might be acquired by means of the exercise of more than one sense modality. Squares are both seen as and felt as square. Rough surfaces are both felt as and seen as rough. Loud sounds are both heard as loud and capable of being identified tactually as loud (as when, for example, one detects the loudness of a drumbeat by the touch of the vibrating surface of the drum). Hot surfaces are capable of being both felt as hot and seen as hot (as when, for example, a blacksmith views a red-hot iron rod).

Of course, one might question whether these concepts are genuinely cross-modal. Whether they are depends not only on whether certain modes of possessing them (say, haptically, or visually) are essential to their possession, but also on whether the experiences that are required for their possession are ones that individuate those concepts. In order for the PC strategy to succeed, it must be seen as making a point, not just about what is essential to
possession of a given concept, but about concept individuation. The strategy moves, from claims about requirements on concept possession (e.g., experiences of red vs. descriptive information about the property red), to claims about concept individuation (e.g., the phenomenal concept red_v vs. the non-phenomenal concept red). This is why Molyneux’s question has such an important bearing on it.

If this is right, the onus falls on the proponent of the PC strategy to argue that, unlike possession of concepts such as square, possession of the phenomenal concept red requires undergoing not merely visual experiences, but visual experiences of the colour red. An initial response to this might be to ask why this might be so. Even if red is modality-specific, and specific to the visual sense modality, that cannot suffice to explain why Mary fails to have this concept. The PC proponent must argue that phenomenal concepts require even finer individuation and possession conditions than modality-specific concepts. She must insist that the phenomenal concept red cannot be possessed, say, by visually accessing readings from a spectrometer and combining this with knowledge that various colours correlate systematically with certain wave reflectance frequencies. It can only be possessed in virtue of experiencing visually the colour red.17

What is wrong with this way of defending the PC strategy is that it fails to distinguish between a claim about how the senses are individuated—a claim which, though controversial, has been thought by many to be true—and a claim about how certain kinds of concepts typically associated with those modalities are individuated—one which is much more controversial and is, I maintain, false.

A sensory mode is a channel or an avenue by which information about objects or phenomena in the world around an organism is transmitted to it.18 A common view is that what individuates the senses is the sets of experiences that are uniquely associated with them. Thus, for example, Grice, when discussing various criteria for distinguishing the senses says,

It might be suggested that two senses, for example seeing and smelling, are to be distinguished by the special introspectible character of the experiences of seeing and smelling; that is, disregarding the differences between the characteristics we learn about by sight and smell, we are entitled to say that seeing is itself different in character from smelling. (1989: 250)

Grice himself argues that this criterion fails to individuate the senses because it is interdependent with another criterion—an Aristotelian one, which individuates the senses in terms of their proper objects (as colours might be thought to do for vision).19

However, even if we were to agree that what individuates different sense modalities as the modalities they are are the experiences uniquely associated with them, it would not follow that the individuation of concepts typically
associated with those sense modalities is uniquely tied to such experiences. This would be to confuse sensory mode with content. Certain types of experiences may uniquely determine a mode to be the sensory mode it is, but content is another matter—it is the information that is transmitted through a sensory mode. And concepts are naturally associated with information—with what a sensory channel delivers to a system—not with the means by which a system might deliver it. It is possible that the qualitative experiences by which we typically come to know that an apple visually present is red—experiences such as the experience of red—are both unique to the visual sense modality and individuate that modality as the sensory mode it is. But it doesn’t follow that the information delivered by that sense modality, and with it, the concept, red, and the concept, experience of red, determine modality-specific concepts, tied to the visual sense.

Further, there are good reasons to reject the claim about concepts, and accordingly, the PC strategy. Return to the situation with Mary and her congenitally blind friend, Harry. The strategy cannot account for the enormous amount of information about colours that Mary and Harry share, in virtue of undergoing experiences tied to sense modalities they each possess and have exercised, and in virtue of concepts possessed thereby. Both Mary and Harry can know that different colours are correlated with distinct wave reflectance frequencies, for example. But they also share much more mundane knowledge about colours. They know that, for example, colours are features of the surfaces of objects and are sensitive to their boundary conditions. More specifically, they both know that objects such as apples come in green and red varieties, that red apples generally taste sweeter than green ones, that limes may be green or orange, and so on. There is an endless amount of information about colours that Harry and Mary share.

The information that Mary and Harry share, it is plausible to think, might suffice for them to possess a concept red. Indeed, proponents of the PC strategy do not wish to deny that a person blind from birth fails to possess any concept of red. So presumably they would not wish to deny that Mary, as well as Harry, possess a non-phenomenal concept of red. Do they have the same concept or not? Well, Mary’s information, unlike Harry’s, will include information acquired by means of the exercise of her visual sense. So do they each have different, non-phenomenal concepts of red? How are we to answer this question?

Suppose that Mary and Harry have been communicating about objects in their environment such as green and red tomatoes, and whether green ones are better for frying. One day Mary is released from her room. When she is presented with a ripe tomato and told it is a ripe tomato, what will she say? Our instinctive response is that she will say ‘Oh, that’s what red looks like!’ Not: ‘Oh, that’s what red is’; not: ‘Oh, now I know what it is to have the phenomenal concept red!’ She speaks to Harry, enthusiastically, about what red looks like. She may have some difficulty communicating here, since after
all she always knew what it was for something to look like something—she has always been able to see. She knows, for example, what square shapes look like, whereas Harry only knows what they feel like. But, given our discussion of Molyneux’s question, we have reason to claim Harry and Mary share the concept, square. And we have reason to claim that they share the non-phenomenal concept red. So perhaps Mary can give Harry some idea of what red looks like by using these, and other, shared concepts.

But how can Mary speak to Harry about what red looks like, if Mary’s phenomenal concept red is a distinct concept from the one that she shares with Harry? It seems that she cannot. What we are envisaging here is something like the situation that Molyneux’s blind subject is in when he first acquires his sight, on the assumption that the answer to Molyneux’s question is ‘no’. That answer supposes that the newly sighted subject might be working with two sets of modality-specific concepts acquired by the exercise of two distinct sense modalities. Similarly, this answer supposes that Mary might be working with two distinct concepts, one of which is acquired by visual experience of the colour red, and is modality-specific, and the other of which is cross-modal and not tied to any particular sense modality.

However, this answer no more plausible than is the negative answer to Molyneux’s question with regard to squares. Just as it is much more plausible to suppose that Molyneux’s subject, when newly possessed of his sight, learns how to extend his concept square, acquired on the basis of tactile experiences, to items visually experienced, it is much more plausible to suppose that Mary, when newly exposed to ripe tomatoes, learns how to extend her concept red, acquired on the basis of information possessed by means other than by visual exposure to the colour red, to visually experienced ripe tomatoes. Just as it is more plausible to suppose that Molyneux’s subject, when newly possessed of his sight, says ‘Oh, that’s what squares look like!’ (Not: ‘Oh, that’s what squares are’, not: ‘Oh, now I know what it is to have the concept square!’), it is more plausible to suppose that Mary, when she is presented with a ripe tomato, will say ‘Oh, that’s what red looks like!’ and that Harry, when he is suddenly given his sight, will say, ‘Oh, that’s what red looks like!’.

If this is right, then more is at stake in the PC strategy than the question-able move from the claim that the senses are individuated by sets of experiences that are uniquely associated with them to the claim that the contents of those experiences are modality-specific. The PC strategy compromises the communicative function of concepts themselves. If concepts are what are shared in communication, the PC strategy cannot be a satisfactory way of attempting to explain the appearance of an explanatory gap.

4. An Alternative

My aim has been to sketch a line of argument that exposes both the empirical and the philosophical weaknesses of a well-known version of the
PC strategy unfavorably in the light of discussion of Molyneux’s strategy. If *a posteriori* physicalism is committed to identities between mental and physical kinds, then, it needs to give an alternative account of the explanatory gap problem. In the light of the preceding arguments, here is what I think is the best strategy for physicalists to pursue.

Mary in her black and white room, and Harry, blind from birth, do possess the concept *red*; a concept that is not tied to the visual sense modality, let alone some specific visual mode of possession. Nevertheless, the problem of the explanatory gap is a conceptual one, and a conceptual one alone. The reason is that many concepts have possession conditions that are disjunctive as across different sense modalities, and disjunctive even within the same sense modality. Spatial concepts such as *square*, *circle* and the like are clear examples of this. But there are other examples, which are perhaps not so obviously spatial, and *red* is one of them. Such concepts may be possessed in more than one way—they may have more than one mode of possession associated with them. And each one of these may be sufficient, although not, by itself, necessary, for their possession. So two individuals, one with sight but no sense of touch, the other with the haptic sense but no sense of sight, can possess the same concept *square*, in virtue of each satisfying one disjunct of a disjunctive possession condition on that concept. Similarly, two individuals can possess the very same concept *red*, in virtue of each satisfying one disjunct of a disjunctive possession condition on that concept. The sighted individual may possess it in virtue of undergoing visual experiences of red, caused by red things in her environment. The blind individual may possess it in virtue of undergoing tactile experiences, reading in braille that red is a surface colour with a certain wave reflective frequency that is sensitive to the boundary conditions of objects. Thus, if a direct recognitional concept is a concept that *must* be possessed in some modality-specific way, then *red* isn’t a direct recognitional concept. Nor is the concept *experience of red*.20

To say this is not to claim that just as concepts are ‘modes of presentation’ of properties or objects, concepts themselves have ‘modes of presentation’; and it is not to appeal to this latter claim to solve the problem of the explanatory gap. The strategy is consistent with a direct reference theory of phenomenal terms or concepts, which rejects ‘modes of presentation’ altogether.21 In committing itself to the view that phenomenal concepts have disjunctive possession conditions which may involve more than one sense modality, then, it is not endorsing the view that such conditions constitute ‘modes of presentation’ of phenomenal concepts. Modes of possession are not modes of presentation.

According to the strategy, the different modes of possession that may comprise a disjunctive possession condition for a concept are not *concepts* at all, not even concepts of concepts. They are *conceptions*—means we use to identify and reidentify their objects as the same things again.22 People can differ quite radically in their conceptions of properties and objects, and still
possess the same concepts of these things, since what makes a concept the concept it is has to do, not with the conceptions that enable us to use it, but with what its function is to identify. What a concept’s function to identify is, is not the same thing as how possessors of that concept are enabled to carry out that function. There are many different ways in which to identify square shapes, no one of which is essential to the concept square. The function of that concept—what individuates it as the concept it is—is to enable its users to identify and reidentify square shapes, to effect reidentification of square shapes under varying circumstances.

How does the distinction between concepts and conceptions, and with it the distinction between individuation conditions and possession conditions of concepts, help to solve the explanatory gap problem? If what I have argued in section 3 is correct, it is plausible to suppose that Molyneux’s subject possesses the concept square on the basis of tactile experiences, and subsequently extends that concept to squares experienced visually. Similarly, it is plausible to suppose that Mary possesses the concept red on the basis of visual and other experiences, other than by the visual experience of red. She extends that concept to red things experienced visually subsequently when she is released from her room. What she acquires on release from her room is not knowledge of a new fact about the world around her. What she acquires is another mode of possession of her concept red.

How does this help us to see why one might wonder how it could be that Brain State B is the experience of red? According to the present suggestion, the concept Brain State B applies to the very same state as does the concept, experience of red. Further, it is plausible to maintain, the concept Brain State B is the very same concept as the concept, experience of red, since both have the function of identifying and reidentifying the state, Brain State B/experience of red. The appearance of an explanatory gap persists because of the mechanisms—conceptions—by which their users are enabled to carry out that function. They cannot reidentify that state without having a conception of it. But this is because they cannot possess a concept of that state without having a conception of it. The conceptions associated with that concept form disjunctive possession conditions on it.

The conceptions associated with the terms ‘Brain State B’ and ‘The experience of red’ are very different. So, even knowing that Brain State B is the experience of red, one might wonder how the conception by means of which one employs the concept, experience of red, could be a conception of Brain State B. That is not the means by which one tracks Brain State B, the mechanism that enables one to reidentify it.

5. Objections

I have suggested an alternative way for a posteriori physicalists committed to identities between mental and physical kinds to deal with the explanatory
gap problem. It is to treat it as one concerning, not the individuation conditions of phenomenal concepts, but their possession conditions, where these are to be understood as providing conceptions of objects and properties of which those concepts are concepts. I would like to conclude by briefly considering three obvious reactions to it.

First, it may seem that the suggestion is simply a stylistic variant of the PC strategy, one that re-locates all of the work that that strategy took concepts to be doing in the conceptions associated with them. Since conceptions are now doing all of the work that concepts did in the PC strategy, concepts look like idle explanatory wheels.

I think this reaction is misplaced. It underestimates the role that (PC) plays in the strategy and so fails to appreciate the distinctive work that concepts and conceptions are doing in the suggested alternative. The claim that phenomenal concepts are modality-specific in the sense of (PC) is critical to the strategy. The suggested alternative rejects the view that there are such things as ‘direct recognitional concepts’, in the sense in which the PC strategy uses it to solve the explanatory gap problem. Conceptions may be modality-specific, but concepts are not. And conceptions are not what one communicates with in interpersonal communication. They are the mechanisms by which one acquires concepts and communicates by means of them.

The suggestion recommends a radical externalism with regard to concept individuation, and with it a coarse-grained view of concepts. According to it, concepts work to ‘hook’ on to objects and properties directly. These latter are the things that help to individuate concepts, whose function is to identify and reidentify them. Conceptions can be, and often are, idiosyncratic as between different speakers. Concepts are not. Crucially, concepts are what one shares with others in communication: they are the carriers of information, and they are the objective basis on which communication about objects and properties depends. Even in the ordinary, or standard case, where two people whose sensory modalities function normally might be talking about red, they might nevertheless have different conceptions of that property. Such differences are typically not what are in play in communication. Conceptions provide us with ways in which to employ concepts to track objects and properties. But they are defeasible ways; they can lead to confusions between concepts. Specifically, they can lead us to suppose that two concepts are in play when only one is. And this can help to explain why one can wonder how it could be that Brain State B is the experience of red.

In order for the explanation to work, both conceptions and concepts must play an explanatory role, where conceptions are not bearers of information.

A second, related objection to the alternative account arises from more general concerns about content. The claim that such terms as ‘Brain State B’ and ‘The experience of red’ express the very same concept is apt to invite objections parallel to those that are voiced against direct reference theories
of content. How could it be that *Brain State B* and *The experience of red* are the very same concept, and one not know it? Consider the, by now familiar, example of Pierre. Living in Paris, he reads about the city he refers to as ‘Londres’ and acquires the belief that he would express with the words ‘Londres est jolie’. Later, when he moves to London and learns to speak English, he acquires the belief that he would express in English as ‘London is ugly’. Assuming that ‘London’ translates ‘Londres’, it seems that ‘Pierre believes that London is pretty’ is true and that ‘Pierre believes that London is pretty’ is false. But how could this be, if ‘Londres’ and ‘London’ express the very same concept?26

This poses a problem for Fregean theorists of content, since the Fregean criterion for substitution of expressions in belief contexts within the scope of the propositional attitude is to require synonymy of substituted expressions. Here is an example where it is evident that that requirement is met, yet substitution fails to preserve the truth-value of the sentences in which the substitution is made. One way to attempt to deal with the problem within the Fregean framework, tracing differences in attitude to differences in cognitive content, is to introduce ‘modes of presentation’ of Fregean senses. However, since the problem can be generated for these modes of presentation too, it seems that this strategy generates an infinite regress of modes of presentation, or senses, the blocking of which at any particular level is semantically arbitrary and ad hoc.

The direct reference theorist’s way of dealing with this puzzle, however, is very different.27 It is to concede that the terms in such cases are synonymous—they do express the same concept—but to claim that such substitutions fail to preserve the truth values of sentences in which they are made because of an additional feature that comes into play in such contexts. This additional feature is the vehicle by which content is expressed. In the case under consideration, the vehicle is the linguistic expression itself. Any alteration in the vehicle by which the content of Pierre’s belief is expressed may alter Pierre’s behavior, not because of a change of content, but because of a change in vehicle. For Pierre might recognize the content of his belief under one name and not under another.

I think that this way of dealing with Pierre-type cases is the right one to pursue in defense of the alternative account. It is true that it is problematic, not least because it openly embraces the view that differences in intentional behavior may result from differences, not in the contents of one’s attitudes, but in syntactic features of their mode of expression. However, my purpose here is not to defend the view in detail. What I hope to have done is to indicate a plausible line of defense that needs to be pursued by any *a posteriori* physicalist who wishes to exploit the alternative strategy for explaining the appearance of an explanatory gap.

A final objection remains to be considered. I’ve argued that that the distinction between concepts and modes of possession of concepts (or
conceptions) can help us to see why Mary doesn’t learn a new fact about the world when she is released from her black and white room and visually experiences red. My claim is that what Mary acquires on release from her room is not knowledge of a new fact about the world around her. What she acquires is another mode of possession of a concept that she possessed before her release from her room, the concept red, where a mode of possession is a means by which one employs a concept to identify and reidentify an object or property in the world one inhabits. But it might be thought that a newly acquired mode of possession is something “newly learned”, and that consequently what Mary acquires on release from her room is a newly learned fact about the world, perhaps a fact about how the property red can be accessed.

There are, in fact, two ways of voicing the worry implicit in this objection that need to be disentangled from one another. The first, which is more easily dealt with, concerns whether acquiring a new mode of possession of a concept counts as acquiring new knowledge of a fact in the world around one, in this case a fact about the property red. The appropriate response to this worry is to say that no new factual knowledge is acquired. What Mary acquires on release from her room is a new way of employing her concept red to access the property, red. The way is through her sense of sight. No new fact about red is thereby acquired, although a new avenue or channel to that property is opened. A mode of possession is a channel to information, in this case, to the property red. Concepts enable their users to identify and reidentify properties and objects, but only through channels. According to the present suggestion, there is more than one channel to the property red by means of which the concept red can be used to identify and reidentify that property. To put the point another way: there is no distinction, corresponding to the distinction between the concept red and a visual mode of possession of it, between the property red and a property of that property. One does not access the property red by means of accessing a property of that property. A visual mode of possession of the property red is not a way of accessing that property by accessing a property of that property, and so, in acquiring that mode of possession one does not thereby learn a new fact about that property.

However, there is another way of pressing the worry expressed above. Suppose that Mary acquires a visual mode of possession of her concept red on release from her room. And suppose that she reflects on what has happened to her. It is natural to say that, on reflection, she comes to realize that there is another channel by means of which she is now able to employ the concept red to access the property red. That is, she comes to realize that she can access a property in the world around her by means of a visual sensory channel. Is this not learning a new fact about the world?

It is tempting to insist here that this is not really a case of acquiring new knowledge about the world, but is, rather, a case of acquiring new knowledge
about how the concept *red* can be exercised or applied to red things. And I think that this would be a legitimate response to make. But it is unlikely to convince someone who persists in thinking that reflection on newly acquired modes of possession reveals new physical facts whose existence shows that physicalism is false. To address the worry fully would take me well beyond the scope of the present paper. But let me attempt to allay the doubts of someone who thinks this by reminding them of the dialectic of the argument between the physicalist and the dualist, in an effort to reveal more precisely what this kind of objection does and does not show about the truth or falsity of physicalism.

According to physicalism, the physical facts are all the facts. The view is taken to imply that if one knows all the physical facts, one knows all the facts. This is a basic premise in the knowledge argument against the view. According to the second version of the present objection, this is false because when Mary is released from her room and *reflects* on her newly acquired mode of possession of her concept *red*, she learns a new fact about channels to information, namely that there is a new visual channel to the property red. The conclusion is that physicalism is false and dualism true.

However, this worry is too general to constitute an objection to physicalism, since it also arises with dualism. According to dualism, the physical and mental facts are all the facts. The view can be taken to imply that if one knows all the physical and mental facts, then one knows all the facts. Now, take all the facts, mental and physical, that there are; and suppose that I can think of all of them. When I think of all the facts, there is a new fact that isn’t included in the set of all the facts I am thinking of. On reflection, this is something I can come to know. But then there is a fact I can come to know, when I reflect on all the facts, that isn’t included in the set of all those facts. Conclusion: dualism is false.

If this objection works, then, it works equally for physicalism and for dualism. It is neutral as between the two ontologies associated with these positions. It cannot therefore be an objection specifically to physicalism. What it shows, I think, is that there is a kind of completeness problem inherent in the formulation of both physicalism and dualism, a problem that is not due to their specific ontological commitments. Since it is not, it is not a problem that the knowledge argument was intended to reveal, and so not one that a physicalist committed to the strategy envisaged here needs to overcome in order to make that strategy a plausible one to pursue.28

6. Conclusion

I have argued that one well-known strategy for defending *a posteriori* physicalism against the explanatory gap problem fails. The strategy appeals to the claim that the problem arises because of the very distinctive nature of phenomenal concepts and so is a conceptual problem alone, having no implications for the physicalist view that the physical facts are all the
facts. My argument does not attack the claim that the explanatory gap problem is a conceptual one. On the contrary, my own suggested strategy for resolving the explanatory gap problem on behalf of \textit{a posteriori} physicalism (and so in its defense) requires that claim. What my argument does attack is a particular way of developing that strategy. This way appeals to the claim that identity statements linking phenomenal and physical kinds such as ‘Brain State B is identical with the experience of red’ open up a kind of explanatory gap because of the distinctive nature of some of the concepts employed in them. Such concepts, such as the concept \textit{red} and the concept \textit{experience of red}, are phenomenal concepts, distinctive types of recognitional concepts. I have argued that considerations about concept individuation tell against this claim. My alternative strategy shows that there is a way of defending \textit{a posteriori} physicalism against objections based on the appearance of an explanatory gap that does not rely on this claim.

Notes

1 Versions of this paper were delivered to audiences at Rutgers University, Columbia University, and the Queen’s University Belfast. I am grateful to those audiences for comments. I am also indebted to Tim Bayne, John Heil, Bill Lycan, Graham Macdonald, Ruth Millikan, and Mark Sainsbury.

2 For those who think that the challenge cannot be met by physicalists, see Jackson 1982, 1986; Nagel 1974, 1986; Levine 1983, 1988, 2001; McGinn 1989; and Chalmers 1996. For those who think that it can, see Papineau 1994, Lycan 1996, Loar 1990, and Tye 1999, 2000, 2001. Despite their differences, particularly with respect to whether failure to meet the challenge shows that physicalism is false (cf. Jackson and Levine), all parties are agreed that there is at least an appearance of an explanatory gap. It is this, and physicalist attempts to explain it, that is the focus of the present paper.

3 Something along the lines of Jackson’s argument seems to have been anticipated by Russell (1927). In a discussion about the supremacy of physics, he says,

There is, however, one important limitation to this. We need to know in what physical circumstances such-and-such a percept will arise, and we must not neglect the more intimate qualitative knowledge which we possess concerning mental events. There will thus remain a certain sphere which will be outside physics. To take a simple instance: physics might, ideally, be able to predict that at such a time my eye would receive a stimulus of a certain sort; it might be able to trace the physical properties of the resulting events in the eye and the brain, one of which is, in fact, a visual percept; but it could not itself give us the knowledge that one of them is a visual percept. It is obvious that a man who can see knows things which a blind man cannot know; but a blind man can know the whole of physics. Thus the knowledge which other men have and he has not is not part of physics. (1927: 389)

4 The phrase ‘explanatory gap’ is Joe Levine’s (1983). He coins it for the view that any physical account of mental phenomena will be incomplete in that it will fail to account for the ‘what-it-is-like’ aspect of certain such phenomena (phenomenal states, such as sensations and visual experiences). Two types of argument are usually thought to support the explanatory gap claim: the knowledge argument (Jackson 1986) and the conceivability argument (Levine 1983,
1998). I have mentioned the knowledge argument in the text. The conceivability argument moves from the claim that psychophysical identities are, if true, necessarily true, and the claim that it is conceivable that there should be two individuals who are physically indiscernible but phenomenologically discernible, to the conclusion that physicalism is false, since the identity statements licensed by it are not necessarily true. Although Levine (1998, 2001) thinks that the conceivability argument is invalid and hence that the appearance of an explanatory gap does not establish that physicalism is false, he also thinks that the challenge—to make room for the phenomenal character of phenomenal experience—cannot be met by physicalists. He puts this failure down to an epistemic gap between mental and physical concepts. To this extent at least, he is in agreement with those such as Tye and Loar, who view the ‘problem’ of the explanatory gap as a conceptual one. However, they do not see it as posing a problem for physicalists. I confine myself in this paper to the strategy’s response to the knowledge argument, although it is clearly capable of being generalized to the conceivability argument (cf. Hill and McLaughlin 1999).

As I have pointed out in the text, although Loar (1990) develops a version of this strategy, my argument is not directed toward his version. This is principally because he restricts the term ‘phenomenal concept’ to concepts of phenomenal qualities, where the latter are properties of phenomenal experiences. So for him, red isn’t a phenomenal concept. See note 10.

This is a stronger assumption than the assumption that mental phenomena and kinds are either identical with or are constituted entirely by physical phenomena and kinds. However, weakening physicalism makes the appearance of an explanatory gap more, rather than less, evident (see, for example, Johnston 1997).

Stoljar 2000 speaks of the red-feeling sensation rather than the experience of red, whereas Tye 1999, 2000 speaks of the visual experience of red. As will become clear, I want to distinguish concepts like red, whose possession might require the having of particular colour experiences (direct recognitional concepts), from ones like pain, whose possession might require the having of particular sensations (sensory concepts). And I want to reject Tye’s view that the phenomenal concept red is one whose possession requires a visual experience of red. So I want to avoid reference either to sensations or to visual experiences when speaking of colour concepts.

Of course, this doesn’t show that all statements that employ phenomenal concepts are a posteriori because the possession conditions of such concepts require the undergoing of particular experiences. ‘Red-feeling experiences are color experiences’ seems to be a priori, for example, even if we suppose, with Tye, that it employs the phenomenal concept red. The concepts embedded in that statement are connected a priori. But that is not the point of the PC strategy. We have no reason to think that psychophysical identity statements like ‘Brain State B is the experience of red’ do employ concepts that are connected a priori. And the appeal to (PC) helps us to see just why that might be.

See, for example, Hill 1991, 1997, Harman 1990, and Hill and McLaughlin 1999. Some who endorse this way of developing the strategy claim that phenomenal concepts are recognitional ones (e.g., Hill and McLaughlin 1999), whereas others (e.g., Tye 1999, 2000) claim that such concepts are direct recognitional ones. I take the difference between the two here to be terminological. Note that, although I am primarily concerned here with how the PC strategy deals with the knowledge argument, Hill and McLaughlin use the strategy to defend a posteriori physicalism against the conceivability argument. So the (AP)-(PC) strategy is quite general.

So, for example, he says, “A person who is blind from birth or who is always restricted to an environment of things with achromatic colours cannot possess a concept with the requisite role and hence cannot possess the phenomenal concept red.” (1999: 712). He makes this claim explicit in Tye 2001. Cf. also Harman 1990: 44: “I am supposing that the person blind from birth has no such concept; that is, the person has no concept of something’s being red that could be immediately brought into service in visual representations of the environment if the person were suddenly to acquire sight.” Both Tye and Harman take the (phenomenal) concept red to be a phenomenal concept because they take it to be a visual concept. This contrasts with those such as Loar 1990, who restricts the term ‘phenomenal concept’ to concepts of qualities of
phenomenal experiences. He gives as examples the concept, *feels like this*, and the concept, *that feature of visual experience*.

11 I am not here assuming that all phenomenal concepts are sensory concepts. Cf. Hill and McLaughlin: “…When one uses a sensory concept to classify one's own current experiences, the experiences that guide and justify one in applying the concept are always identical with the experiences to which the concept is applied.” (1999: 448). This seems to rule the phenomenal concept *red* out as a sensory concept, since that concept does not apply to the *experiences* that guide and justify one in applying it. However, some *a posteriori* physicalists, like Tye and Harman, would count concepts like *red* as phenomenal, if not specifically sensory, ones. And many philosophers would classify *red* as a recognitional concept, if not a ‘direct’ recognitional one, despite the fact that the experiences required for its possession are not identical with what it applies to (viz. the property red, rather than an experience). Cf. Fodor 1998. I take the physicalist strategy under consideration to be one that treats the concepts *red* and *experience of red* as both direct recognitional and phenomenal, even though it may be that the latter but not the former is, more specifically still, a sensory concept.

12 For an excellent survey of the history of the debate, see Morgan 1977.

13 The original formulation by Molyneux, quoted by Locke, is this:

Suppose a Man born blind, and now adult, and taught by his touch to distinguish between a Cube and a Sphere of the same metal, and nighly of the same bigness, so as to tell, when he felt one and t’other, which is the Cube, which the Sphere. Suppose then the Cube and Sphere placed on a Table, and the Blind Man to be made to see: quaere, Whether by his sight, before he touched them, he could now distinguish, and tell, which is the Globe, which the Cube (Locke, 1975: 146).

Although Locke is concerned with whether the blind man would be able to apply three-dimensional concepts such as *sphere* and *cube* on the basis of vision, it seems clear that the question is whether he would be able to extend his two-dimensional concepts based on haptic experiences, such as *circle* and *square*, to visually identified shapes. See Evans 1985.

14 As Gareth Evans puts the point,

…Molyneux’s Question is about whether a born-blind man who can see a circle and a square would extend his concepts to them. It is not a question about how soon after the operation, and via what process, a newly sighted man would be able to see. Molyneux’s Question requires only that the newly sighted man would be able to have visual experiences of circles and squares without his, or his brain’s, having had a chance to establish correlations between the old and the new information. (1985: 154–6)

For a very useful recent survey of some of the ways in which Molyneux’s Question can be interpreted, see Eilan 1993.

15 Nor need they insist that the undergoing of experiences of both sorts is a necessary condition on possession of such concepts (perhaps a disjunctive requirement would suffice, for example).

16 This is not to say that the experiences themselves are not, or do not have elements that are, tied to specific sense modalities. It may be that the tactile experience of square, or of rough, has elements that are distinctive to the haptic sense modality and are not echoed in experiences acquired through the exercise of other sense modalities, such as sight. But this does not favour a negative answer to Molyneux’s question over an affirmative one, since what answer we give crucially depends on whether the spatial content of such experiences are extrinsic to them. For more on this see Eilan 1993.
The question whether a concept is modality-specific, or 'tied' to a particular sense modality, concerns whether its possession requires the operation of a single sense modality. When I say that a concept is one whose possession may allow the undergoing of experiences tied to more than a single sense modality, I mean that it may allow the exercise of more than one sense modality, each of which delivers up modality-specific experiences. I do not mean that the experiences themselves involve the operation of more than one sense modality (so that, say, tactile experiences may be had through the exercise of the visual sense modality).

This makes the problem here not quite the problem posed by Molyneux's question, but rather a variant of it.

According to the Shorter Oxford English Dictionary, a sense is “Each of the special faculties, connected with a bodily organ, by which man and other animals perceive external objects and changes in the condition of their own bodies. Usually reckoned as five – sight, hearing, smell, taste, touch”.

Also, see Keely 2002, who argues that it is false that the senses are individuated by the qualitative character of perceptual experience.

Nor, according to this strategy, is the concept pain, although the point cannot be developed here.

For a direct reference theory of content, see Salmon 1986.

See Millikan 2000. She takes concepts to be abilities, one such being the ability to identify and re-identify an object or property. She characterizes this ability in functional terms, in accordance with the theory articulated in detail in Millikan 1984.

Although I am appealing to Millikan’s account of concepts here on behalf of the a posteriori physicalist, the concept of function that is required by this strategy is neutral as between her own, historically-based account and a causal-role account (such as Fodor’s 1990).

Cf. Millikan: “the mechanisms of perceptual constancy that enable us to perceive, for example, the same color, shape, voice, or moving object as being the same one through diverse proximal stimulations, diverse intervening media, and various kinds of distortions and static, exemplify our ability to make the same perceptual judgment in a variety of ways. So does our ability to use different senses to confirm the same judgment perceptually. Given a variety of ways to observe the same state of affairs, none of these methods is definitional of the concepts employed . . . .” (2000: 103)

This represents a slight departure from Millikan’s view of the matter. According to her, in one sense, to say that two people have the same concept is just to say that they have concepts of the same object or property in the world. (This is the sense in which I am saying that Brain State B and the experience of red are the same concept.) But she goes on to say that in another sense, a single person can have two ‘detached’ concepts of the same object or property, and this happens when her concepts are individuated, not by their content, but by the conceptions associated with them. Since her official view is that concepts are not to be confused with conceptions and that concepts are individuated by the objects or properties of which they are concepts, I here ignore this complication in her view.

For other examples of this kind and the difficulties they pose for Fregean theories of content, see Mates’ classic 1952.


I am indebted to Graham Macdonald for discussion on this point.

References


———. 1986 The View From Nowhere, Oxford: Oxford University Press.


