Physicalism and Phenomenal Consciousness

Robert Kirk

I. INTRODUCTION

Most of us are happy to suppose that stars, planets, shoes, ships, and everything else we think of as inanimate, are composed of nothing but the fundamental particles of physics. Physicalism, roughly, is the view that not only all inanimate things, but we too, are composed of nothing but those basic, purely physical items. An influential line of argument against physicalism appeals to ‘Cartesian intuitions’ which suggest that the world might have been physically just as physicalists suppose it is, but mentally different. The ‘transposed qualia’ intuition is that the subjective character of our experiences in some sensory modality (perhaps colour vision) might have been systematically swapped round relative to how we actually experience the world,
while the physical facts were as physicalists suppose them to be. The ‘absent qualia’ intuition is that the world might have been physically as physicalists suppose but without phenomenal consciousness: a ‘zombie world’.1 The argument based on these intuitions goes:

(1) If physicalism is true, transposed and absent qualia are impossible.
(2) Transposed and absent qualia are possible, or at least cannot be ruled out a priori.
(3) Therefore physicalism is false.2

The argument is valid, but in order to assess its premisses we need to be reasonably clear about what sort of possibility is involved and what physicalism commits you to; for there are many misconceptions and confusions in this area. I think the biggest and most influential misconceptions are the Cartesian intuitions themselves. I agree with the view that they depend on a radically wrong conception of phenomenal consciousness and can be ruled out a priori, so that premiss (2) is false and the above argument fails for that reason. However, I have defended that view elsewhere and will not rehearse my argument here [Kirk (2005), pp. 37-57; 2008].

Here I have two main aims: one is to make clear that physicalism commits you to a strong version of premiss (1), where the relevant sense of ‘possible’ is something like ‘not inconsistent or incoherent for broadly logical or conceptual reasons’. The other is to correct the misconceptions that have led some physicalists to suppose they are not really committed to premiss (1), together with a different lot of misconceptions (other than the Cartesian intuitions themselves) which suggest that this commitment is untenable. It will be important to distinguish between the claim that physicalism requires the physical facts to entail the phenomenal facts by logico-conceptual necessity (on which I will say more in sections IV and VI), and the claim that it requires the phenomenal facts to be knowable a priori on the basis of the physical facts. While physicalism commits you to the first claim, it does not commit you to the second.

II. PHYSICALISM, REDescription, AND STRICT IMPLICATION

The difficulties facing physicalism arise in connection with any theory which performs anything like the functions of today’s physics, assuming it does not invoke consciousness or intentionality; the theory’s details are therefore irrelevant. That means we can pretend we have at our disposal a true physics which does not directly invoke consciousness, intentionality, or other psychological notions, and in terms of which it is possible to explain the workings of physical systems.
Imagine, then, that P is the conjunction of all true factual statements about the universe in terms of that physics, including – vitally – all physical laws. Then we can take it that P is a true description of everything physical in the whole of space and time. Of course there are vastly more truths than those austerely physical ones. Some are expressible in ordinary language, others in the vocabularies of the special sciences. It seems clear that, for physicalists, those other truths are different ways of describing the world described by P. But they are not just redescriptions: they are pure redescriptions in the following sense. Consider the statement, ‘A man jumped out of a bath shouting’. One ‘pure’ redescription of an aspect of that state of affairs would be, ‘A man jumped’; another would be ‘a bath existed’. A non-pure redescription would be, ‘Archimedes shouted “Eureka”’. The truth of a pure redescription of an item relative to a given description depends only on what the latter entails, independently of the item’s relations to other things.

The redescription thesis is:

\[(R) \text{ All truths about the world not included in P are (with an exception to be explained shortly) pure redescriptions of the world specified by P.} \]

It seems that all physicalists are committed to (R). To reject it would imply there were truths about the actual world that would not be made true purely by a world which satisfied P. In that case something not included in the world specified by P would have been needed to account for those truths. I doubt if any physicalists would find (R) objectionable, although it does need further clarification. The main part of clarifying it, I suggest, consists in pointing out that it commits you to the ‘strict implication thesis’, understood as follows. I take it that a statement A strictly implies a statement B just in case ‘A and not B’ is inconsistent or otherwise incoherent for logical or conceptual reasons. The thesis is:

\[P \text{ strictly implies Q.} \]

As before, P is the conjunction of all austerely physical truths; Q is the conjunction of all interesting other truths. (This category of ‘interesting’ truths avoids – I think harmlessly – complications arising from statements such as ‘Socrates does not have an immaterial soul’ or ‘There are no angels’.) However, the following interesting and important statement will be endorsed by all physicalists, yet is not strictly implied by P:

\[(N) \text{ Nothing exists other than what is strictly implied to exist by P.} \]
To reject (N) would be to concede that dualism was true. For although the world explicitly specified by P is purely physical, P by itself appears to be compatible with the existence of non-physical items such as angels.

III. WHY PHYSICALISM COMMITS YOU TO THE STRICT IMPLICATION THESIS

Suppose physicalists were to deny the strict implication thesis. That would commit them to there being a purely physical possible world w where P was true and Q false; w would be in all physical respects just as physicalists think our world is, but somehow different. But that is inconsistent with physicalism, as the following argument shows.

1. Assume, for reduction, that there is a possible world w where (a) P is true and (N) holds (that is, nothing exists other than what is strictly implied by P); (b) Q is false.
2. (b) entails there is a difference between w and our world, a difference which can consist only of the presence of something physical or non-physical either in w or in our world. (Recall that the same physical laws hold in both worlds.)
3. There is no physical difference between our world and w [by 1 and the definition of P].
4. There is nothing non-physical in w [by 1 and the definition of P].
5. Therefore there is something non-physical in our world, in which case physicalism is false [2, 3, 4].
6. Therefore physicalism cannot consistently allow that w is so much as logically or conceptually possible: physicalists must reject assumption (1) [1, 5].
7. Therefore physicalism commits you to the strict implication thesis [6, definition of strict implication].

You might object that the conclusion conflicts with the widespread assumption that physicalism is only contingently true. Strict implication holds by logical or conceptual necessity; so if physicalists are committed to the strict implication thesis, they seem committed to physicalism itself being true by logical or conceptual necessity. However, the thesis is ambiguous. Suppose I arbitrarily select a book and pick out the first statement on page 1 and the first statement on page 2. On inspecting them I find the first strictly implies the second. Of course it is a logical or conceptual matter whether that relation holds between those two particular statements. On the other hand it is contingent just which statements are the first on page 1 and the first on page 2. So the following statement:
(S) The first statement strictly implies the second

may be construed in either of two ways: (i) so that the expressions ‘the first
statement’ and ‘the second’ each picks out a particular statement rigidly; (ii)
so that those expressions are taken to mean respectively ‘whatever the first
statement may happen to be’ and ‘whatever the second may happen to be’.
Construed the first way, (S) is either necessarily true or necessarily false.
Construed the second way it is contingent, because its truth value depends on
which statements those descriptions happen to pick out.

Similarly, the strict implication thesis is construable in either of two
ways. If we construe it so that ‘P’ and ‘Q’ each picks out a particular con-
junction of statements rigidly, it is necessarily true or necessarily false. But
the second construal is also available – and both more natural and more to
the point. It is not as if we actually knew just which purely physical truths hold
or which interesting other truths hold; the point of physicalism is expressed
by saying that regardless of what they may be in detail, the former strictly
imply the latter. Each possible world is such that either the purely physical
statements true in it strictly imply the other interesting truths, or they do not;
possibly our world belongs to the former class, possibly it belongs to the lat-
ter. By saying physicalists are committed to the strict implication thesis, I
mean they are committed to the thesis that our world belongs to the former
class. In that way the strict implication thesis can be construed – and is here
to be construed – as contingent and empirical.

The strict implication thesis is significantly different from what David
Chalmers and Frank Jackson say physicalists are committed to [Chalmers and
Jackson (2001)]. Chalmers classifies physicalism into two main types, A and
B. Type-A physicalists hold that ‘phenomenal truths (in so far as there are
such truths) are necessitated a priori by physical truths’; type-B physicalists
‘accept that phenomenal truths are not necessitated a priori by physical truths,
but hold that they are necessitated a posteriori by physical truths’ [Chalmers
(1999): pp. 473f.]. He and Jackson say a conditional holds by such necessita-
tion ‘when it is possible to know that P entails Q with justification independ-
ent of experience’ [(2001): p. 316] (so that an a posteriori entailment is a
conditional which holds when that is not possible). They have argued that
physicalists are committed to the former – the ‘a priori thesis’: that phenome-
nal truths are necessitated a priori by physical truths. I agree broadly with
their reasoning [Kirk (1974); (2005)]. However, their epistemological version
of the connection from physical to mental is significantly different from the
ontological one expressed by the strict implication thesis. As I will explain
later, there is a clear sense in which physicalists can endorse the strict impli-
cation thesis while rejecting the ‘a priori thesis’; misconceptions can easily
arise from ignoring this distinction (see sections V and VI).
An obvious and important consequence of the fact that physicalism commits you to the strict implication thesis is that ‘a posteriori’ or ‘type-B’ physicalism, according to which phenomenal truths are not necessitated by physical truths a priori, but only a posteriori, is not physicalism at all because it rules out the strict implication thesis. The assumption that it is on the contrary a tenable variety of physicalism turns out to be a significant misconception.

Some philosophers believe physicalists can escape the problems posed by the Cartesian intuitions simply by endorsing the psycho-physical identity thesis. But there seems no good reason why they should adopt the identity thesis – given they are committed to the strict implication thesis. The strict implication thesis is compulsory for physicalists, the identity thesis is not.

IV. SUPERFICIALLY ‘ANALYTIC’ LINKS ARE NOT REQUIRED

An apparent difficulty for exponents of the strict implication thesis is that it seems to require analytic links from P to Q. How otherwise could the implication hold for ‘broadly logical or conceptual reasons’? And the Cartesian intuitions look like powerful reasons to reject the view that there are such links, at least if ‘analytic’ links are determined by the meanings of the sentences involved. For the meanings of the austerely physical truths in P are very different from those of the mental truths in Q, especially those dealing with phenomenal consciousness. It may seem that either physicalism doesn’t after all commit you to the strict implication thesis, or else physicalism is false.

Again the apparent difficulty arises from mistaken assumptions. Consider for example Christopher Hill’s assertion that ‘there are no substantive a priori ties between the concept of pain and the concept of C-fiber stimulation’, an assertion he supports by saying ‘it is in principle possible to master either of these concepts fully without having mastered the other’ [Hill (1997), p. 76; cf. Papineau (2002), p. 49]. If my earlier reasoning is sound, Hill’s last claim, though true, is beside the point. It is after all possible to master fully the concept of the ratio of a square’s diagonal to its side without having mastered the concept of irrational number; yet what satisfies the first description satisfies the second for broadly logical or conceptual reasons. Or consider on the one hand P, which is formulated exclusively in terms of narrowly physical concepts, and on the other the statement ‘There are rivers’. It is certainly possible (though no doubt unusual) for an individual to have mastered the austerely physical concepts engaged by P without also mastering the concepts involved in ‘There are rivers’. Yet physicalists cannot deny that ‘P and there are no rivers’ is incoherent for broadly logical or conceptual reasons. For the universe specified by P includes at least one planet where quantities of liquid flow along numerous channels in its various land masses, and drain into larger masses of that liquid. To maintain that P might be true even if there were
no rivers would be inconsistent with a grasp of both sets of concepts. We might say there are deep and superficial senses of ‘analytic’ and ‘conceptual connection’. Superficially there are no conceptual or analytic links between austerely physical statements and statements in other terms; but there are deep connections which hold for broadly logical or conceptual reasons. 5

Relatedly, we can come to know that the relation of strict implication holds without having to construct a chain of deductive inference from the antecedent to the consequent. It is enough if, while P specifies the world in its terms, the reality thus specified qualifies for description in the terms in question, which might be radically different. There is no mystery here. Suppose we have a square array of 1,000 x 1,000 pixels, each of which may be either black or white. We can specify a complete assignment of pixels on the lines of ‘(1,1)B, (1,2)W, ... (2,1)B, ... (2,555)W, ... (1000,1)B, ... (1000,1000)W’. The set of concepts involved in that simple code is meagre; yet the whole array might be immediately and unmistakably recognizable as, say, an image of a horse. Given that particular pixel specification – and assuming, as we may, that the image is indeed easily identifiable as that of a horse – it must be for broadly logical or conceptual reasons impossible for the description ‘image of a horse’ to fail to apply. Yet there are no superficial conceptual or analytic links from formulas of the specifying code to ‘image of a horse’; nor is there any string of logical deductions from the former to the latter [Kirk (2005), pp. 17-23].

V. JACKSON’S MARY AND PURE PHENOMENAL CONCEPTS

Mental facts, especially phenomenal ones, may still seem an insuperable obstacle to the tenability of the strict implication thesis. But here again, as we shall see, it is easy to succumb to misconceptions. Consider Frank Jackson’s [1982] example of Mary, a scientist imprisoned from birth in a totally black and white environment, who knows all the physical and functional facts relevant to perception, but not what it is like (for example) to see the blue sky. Many people believe Mary brings out what is special about phenomenal truths, and shows we cannot move a priori from physical facts to phenomenal facts. If that is right, it may seem to threaten the strict implication thesis.

Jackson’s main point was that although Mary has all the relevant physical and functional knowledge, she cannot on its basis alone come to know what it is like to see the various colours. For (he assumes) in order to know what it is like, one must actually have had such experiences or at least to be able to summon them up in imagination. (That assumption seems to fit our grasp of the concepts involved and I take it to be correct.) I will not go over the familiar ground of Jackson’s reasoning, but inquire whether the story of Mary undermines the strict implication thesis: does it show that the physical
facts do not strictly imply the phenomenal facts? Let us consider first whether it shows that Mary cannot move a priori from the relevant physical facts embodied in P to knowledge of truths about colours and colour experiences.

A preliminary point. The a priori thesis would not be interesting if the only reason physicalists could not endorse it were that the necessary concepts could not be acquired a priori from P. We should still want to know whether the a priori thesis would hold if the investigator did already have the concepts, since the guiding question here is whether the physical facts fix the phenomenal facts for broadly logical or conceptual reasons; and the answer to that question does not depend on which concepts people may happen to possess. But now we face a difficulty. As noted just now, I take it that grasping phenomenal concepts includes knowing what it is like to have experiences in the relevant sensory modality (here, knowing what it is like to see colours). But it would be pointless to require Mary, whose task is to discover what it is like to see colours, to come to it already equipped with that knowledge. To avoid trivialization, therefore, we had better require her to start with no more in the way of colour concepts than whatever abstract components of them she can be supposed to have in her grey prison; what she can know about colours without having experienced them.

She can have quite a lot of such abstract knowledge: for example that colours are (taken to be) properties of objects normally easy for normal viewers to detect in normal lighting conditions; that anything visible at all has some colour; that colours can be ordered so as to form a solid whose dimensions are hue, lightness, and intensity; that some colours strike us as more or less similar in each of those respects. She can also learn in what commonly identifiable circumstances – and also in what microphysical circumstances – people are disposed to assent to statements about colours and colour experiences, and which descriptions of colours and colour experiences go with which colour words.

Could she really get all that information a priori from P? I think so, given P is assumed to be the physical truth about the world as physicalists conceive of it. On that assumption it could hardly be denied that P would fix all merely functional and dispositional facts. (Chalmers and Jackson describe a thought experiment designed to establish the same conclusion [Chalmers and Jackson (2001),pp. 334f.].) But now, in the absence of defeating facts – and commonly they are absent – the ordinary phenomenal statements we are disposed to assent to in normal circumstances will actually be true. (By ‘ordinary phenomenal statements’ I mean those expressed in ordinary language and suitable for conveying to others what a certain kind of conscious experience is like: for example, ‘It was like hearing thunder’, or ‘I’ve got a headache’.) I take it this is a priori knowledge.6

On that assumption Mary will be able to move a priori from P to truths about colour experiences – if that requires no more than being able to discover, from knowing P, which such statements are true. And that is indeed all
it requires according to Chalmers and Jackson’s definition of ‘a priori entailment’: they say P a priori entails Q ‘when it is possible to know that P entails Q with justification independent of experience’ [Chalmers and Jackson (2001): p. 316]. In one clear sense, then, Mary can move a priori from P to phenomenal truths. But there is a snag.

Although she can discover a priori from P which statements about colour experiences are true, she does not fully understand those statements; that would require her to have actually had experiences of colour or to be able to conjure them up in imagination. Since knowing P is obviously not enough to confer either such experiences or the necessary imaginative capacity, she cannot learn from P alone what it is like to see colours. Therefore she cannot come to know with full understanding any phenomenal truths about colour experiences a priori from P. The same goes for others who start off with no more than Mary’s ‘abstract’ knowledge of the nature of colour concepts. So now we have two senses in which P might a priori entail ordinary phenomenal truths. The first (Chalmers and Jackson’s) does not require the investigator fully to understand the truths in question; the second does. In the first sense phenomenal truths are a priori entailed by P; in the second they are not. Notice, however, that their being entailed in the first sense entails in turn that they are strictly implied by P. For if these truths are discoverable a priori from P, then there is no logico-conceptually possible world where P is true and any of them is false, which means they are strictly implied by P. This demonstrates the importance of distinguishing the strict implication thesis from the a priori thesis. (It incidentally offers its own special route to showing that the Mary example does not refute physicalism.)

However, that phenomenal truths are not a priori entailed in the second sense constitutes a significant ‘explanatory gap’, which physicalists must acknowledge and explain or explain away [see Levine (1983) and the next section]. This gap is thought to be particularly problematic for the case of ‘pure’ phenomenal truths. Suppose for example that Mary, still in her grey prison, is able to conjure up in imagination experiences like the ones she would have had if she were actually seeing colours. If she picks out one of these experiences and calls it ‘r’, she can say things like:

(T) I am having an experience of r.

The concept engaged here is a pure phenomenal concept: a concept one can apply to an experience of one’s own without having to know any particular physical, behavioural, or functional facts about it. Such concepts are not necessarily linked to any verbal dispositions that others could understand, so Mary’s utterances of (T) might well leave us completely in the dark about the character of her experience. Such pure phenomenal truths therefore seem to float free from the physical facts. Again, however, by the reasoning sketched
earlier, she (and for that matter others) could still work out a priori from P that she is normally disposed to utter (T) when such and such a brain process (say b) occurs, hence, if she is in fact so disposed on this occasion, that (T) is true. However, neither she nor anyone else could also work out a priori from P that she was having an experience of the particular kind she has named ‘r’. The reason is that, as for ordinary phenomenal concepts, P by itself does not enable her or anyone else to connect the relevant P-information with the character of her experience r; she can only make that connection empirically, when she actually has the experience. Pure phenomenal truths make the gap between physical and phenomenal facts look dauntingly wide.

VI. THE EXPLANATORY GAP AND ‘TRANSPARENCY’

How can physical, behavioural, and functional facts be ‘logically or conceptually’ linked to pure phenomenal truths? To answer this question physicalists can appeal to the distinction between superficial and deep connections, and argue that these truths do not pose an additional problem, on top of that of explaining the nature of phenomenal consciousness itself.

The lack of superficial connections should not be surprising. Part of being phenomenally conscious is that we can apply concepts to our experiences just because we have them, without needing to learn what causes them, what physical processes underlie them, or what behavioural dispositions they may activate. That explains why our experiences appear to be independent of the physical facts – but obviously does not entail that they are. (Carruthers makes a similar point [Carruthers (2005), p. 78].) If physicalists can explain how there is phenomenal consciousness, therefore, they will have no difficulty in explaining how we can describe our experiences in ways apparently independent of the physical facts. Still, can there really be logico-conceptual links from physical facts to pure phenomenal truths?

The problem is often stated in terms of ‘transparency’. Thomas Nagel, for example, has argued that we need a theory that ‘would render transparent the relation between mental and physical’ [Nagel (1998), p. 351. Cf. Chalmers (2007), p. 174; Levine (1993)]. Colin McGinn, like Joseph Levine, thinks physicalism requires ‘reduction’ and demands ‘transparent bottom-up necessity from the reducing facts to the reduced facts’, asserting that ‘reduction is possible only in the presence of a priori entailments from one set of facts to the other’ [McGinn (2004), p. 13]. ‘A priori entailments’, he says, are statements of the form “P if and only if Q” that can be known a priori: that is, there has to be a conceptual bridge spanning the biconditional. The concepts have to be a priori connectable in virtue of being the concepts they are [McGinn (2004), p. 19].
McGinn overlooks the kind of connection I classify as strict implication, as becomes clear from what he takes to be ‘a clear obstacle to achieving such a priori connections, namely the vastly different concepts that figure on either side of the biconditional’ [McGinn (2004), p. 19; cf. Papineau (2002), p. 49]. Earlier I argued that, to the contrary, mere difference in concepts – even radical – is no obstacle to strict implication; so I think McGinn’s requirement is too strong.

We still want as much transparency as possible. What can reasonably be required? Nagel and Jackson are right to point out that when considering some particular organism, we cannot move a priori from P to knowledge with full understanding of what it is like for that organism. Such transparency is absolutely unattainable. But that is not a problem for physicalism since, as I have argued, it does not entail that phenomenal truths are not strictly implied by P. And if we take care to avoid another confusion, we shall be able to see how physicalists can consistently maintain that there are deep logico-conceptual links from the physical to the phenomenal.

The threatened confusion comes from a tendency to assume that if we cannot discover a priori from P what it is like for the bat, robot, or other system under consideration, then we cannot discover a priori from P that it is like anything: that the system is phenomenally conscious at all. That is a mistake; the two questions are distinct. Physicalists can accept that the ‘what-is-it-like’ question is unanswerable while offering an answer – perhaps a functionalist one – to the ‘is-it-like-anything’ question (see Kirk (2005), pp. 61-64). A platitude makes this approach workable: no one could be conscious without being so in some particular way; there is no such thing as being phenomenally conscious but having no particular kind of experience when seeing red, for example. It follows that if physicalists can formulate necessary and sufficient conditions for phenomenal consciousness whose satisfaction could be strictly implied by the physical facts – admittedly a big ‘if’ – satisfaction of those conditions will fix not only that there is something it is like for the individual in question, but what it is like. If consciousness itself is strictly implied by P, so are individual phenomenal truths, pure ones included.

A possible objection: even if God could not have created the physical world as it is without thereby ensuring that its inhabitants were conscious in specific ways, couldn’t he have distributed phenomenal properties differently? That is just an appeal to the transposed qualia intuition. If it points to a genuine possibility, then physicalism is false for that reason – by premiss (1) of the anti-physicalist argument stated at the beginning of this paper. I concede that the intuition has some force: it does seem to support premiss (2). However, I am not claiming to disprove the Cartesian intuitions here: only to explain why physicalists must accept premiss (1), properly understood. What matters now is that physicalists can consistently regard the platitude just noted as a licence to omit from their accounts of phenomenal consciousness
any attempt to explain why the character of a given individual’s experiences is as it is. It seems clear that they can take this line even if I was wrong to maintain earlier that Mary could discover a priori from P which phenomenal statements are true, in spite of not being able to understand them on that basis.

If that is right, then it will be for broadly logical or conceptual reasons that the phenomenal truths obtain, given the physical facts. So, provided physicalists can explain how it is that the physical facts strictly imply that human and other organisms are phenomenally conscious at all, they will have provided as much transparency about the relations between the physical and the mental as can consistently be demanded.

VII. HOW THE MAIN POINTS ARE RELATED

The Cartesian intuitions are a crux for anyone aiming to get clear about phenomenal consciousness. I have argued that physicalism commits you to the strict implication thesis, hence to a strong construal of premiss (1) of the anti-physicalist argument. Neither the identity thesis nor ‘a posteriori physicalism’ offers an escape from that commitment, so physicalists must disprove premiss (2), explaining why the states of affairs suggested by the Cartesian intuitions are not even logico-conceptually possible. However, this task is not as hard as some suppose because strict implication does not require analytic links from physical to mental, nor does it require the phenomenal facts to be knowable a priori on the basis of the physical facts. Since the Cartesian intuitions are the main obstacle to functionalist accounts of consciousness, disproving (2) will greatly facilitate the task of explaining what it takes for a system to be phenomenally conscious: explaining that there is ‘something it is like’. If that can be done, there will be no additional need to explain what it is like.10,11
are bound to their subject is no stronger than nomic. That entails there are logico-conceptually possible worlds where P is true but the laws of nature fail to bind any qualia at all to the physical facts: they are absent qualia worlds. See also Kirk (2005), pp. 56f.

2 There are early versions of this argument in Stout (1931); Campbell (1970); Kirk (1974) (who offers mistaken arguments purporting to show that the zombie possibility is genuine); Kripke (1972/80). It is powerfully developed in Chalmers (1996).

3 See Kirk (2005), pp. 12-17. It might be argued that exponents of the strict implication thesis could maintain that because strict implication is not the same as a priori necessitation, phenomenal truths are indeed not necessitated a priori by physical truths. However, it is clear that Chalmers would still classify such physicalists as type-A rather than type-B: he and Jackson seem not to allow a significant difference between the ‘a priori thesis’ and the strict implication thesis.

4 For example Block and Stalnaker (1999). Some think the only route to an adequate explanation of physicalism is via identity: e.g. Levine (1993); McGinn (2004). I claim the arguments in this paper show they are mistaken. See also section VI below.


6 I also take it she could in principle discover a priori from P, at least on some occasions, that there were no defeating facts. However, if it is not a priori that people are normally disposed to tell the truth, then I don’t think phenomenal truths are discoverable a priori from P. That would undermine the present line of reasoning but not also (for a reason to be given in the next section) my eventual conclusion.

7 Some have argued that the best route for physicalists to take round the difficulties posed by the Cartesian intuitions is through a proper grasp of the nature of phenomenal concepts. (That is the ‘phenomenal concept strategy’ attacked by Chalmers in his (2007).) Although such understanding is obviously necessary, what generates the potential for phenomenal concepts and at the same time the difficulties encapsulated in the Cartesian intuitions and the explanatory gap is phenomenal consciousness itself. Grasping the nature of the concepts cannot provide an understanding of that.

8 McGinn’s insistence on biconditionals, rather than merely conditionals from the physical to the mental, is unnecessary if, as argued here, physicalism needs only strict implication rather than identity.

9 See also note 1.

10 These tasks are attempted in Kirk (2005).

11 Thanks to Bill Fish, and to the participants in a seminar at the University of Lancaster, for discussions of the topics covered here.

REFERENCES


--, (2008), ‘The Inconceivability of Zombies’, Philosophical Studies vol. 139, pp. 73-89.


