

Chapter 10

Explanation and two conceptions of the physical

(*Erkenntnis* 62: 71-89, 2005)

Preview

Any position that promises genuine progress on the mind-body problem deserves attention.

Recently, Daniel Stoljar has identified a physicalist version of Russell's notion of neutral monism; he elegantly argues that with this type of physicalism it is possible to disambiguate the notion of physicalism in such a way that the problem is resolved ('Two conceptions of the physical'.

Philosophy and Phenomenological Research LXII: 253–81, 2001; 'The conceivability argument and two conceptions of the physical'. *Nous* 35(s15): 393–413, 2001). The further issue then arises of whether we have reason to believe that this type of physicalism is in fact true. Ultimately, one needs to argue for this position by inference to the best explanation, and I show that this new type of physicalism does not hold promise of more explanatory prowess than its relevant rivals, and that, whether it is better than its rivals or not, it is doubtful whether it would furnish us with genuine explanations of the phenomenal at all.

Chapter 10

Explanation and two conceptions of the physical

1. Introduction.

One can easily conceive how the structural and functional facts represented by the physical truths fall short of giving us the intrinsic facts represented by the qualitative truths concerning our phenomenal or conscious states. For people attracted to physicalism, consciousness therefore appears to be a mystery (Chalmers 1996). In its least committed, epistemological version we can speak of this in terms of the existence of an explanatory gap (Levine 1983) between the physical truths and the qualitative truths: even if we have all the physical truths we may still have no conception of how to explain the constitution of the qualitative truths. This is a deep problem for physicalism about consciousness, but it is not easy to respond by simply rejecting physicalism and adopting a type of dualism instead. Dualism is widely recognised to be problematic too, since it cannot account for mental causation under the assumptions that the physical is causally closed and that there is no general causal overdetermination (e.g., Kim 1998). But if neither physicalism nor dualism seem viable, then our problem—the mind-body problem—is that it appears that we cannot say anything at all about what consciousness is.

I shall examine an intriguing position that promises to resolve this problem. It is a physicalist cousin of Russell's notion of 'neutral monism', put forward recently by Daniel Stoljar (2001a, 2001b).¹

I will grant the general metaphysical framework of this type of position, but I shall argue that it is not as compelling as it initially seems: there is little reason to believe this kind of physicalism is better than its relevant rivals at resolving the problem, and there is reason to doubt that it accounts particularly well for the explanatory gap.

2. Two conceptions of the physical and the mind–body debate.

This section is devoted to presenting and explaining how Stoljar sets up the debate and resolves the problem. Though many of the opening moves are controversial and open to debate, I propose to set aside much of this controversy in order to focus on Stoljar's main contribution.

Two conceptions of the physical

If you are a physicalist, then you had better have a more inclusive conception of the physical than you might think. Physical theory deals in structure and function and thus characterises only extrinsic, dispositional properties. But more properties may need to count as physical. Many people hold that dispositions have intrinsic categorical bases. There are good reasons for this since dispositional properties are causally efficacious only in virtue of their categorical bases. The physicalist must say that these intrinsic properties are physical too. If they are not physical then physicalism is false.

So the physical comprises both the extrinsic, dispositional properties that physical theory tells us about and, in addition, intrinsic, categorical properties that current physical theory is silent about. These two classes of properties are not co-extensive. We can conceive of similar classes of dispositional properties with different classes of intrinsic categorical properties (Stoljar 2001a: 260–1).

Together with Stoljar, let us call the first type of physical properties t-physical properties because they are what physical theory tells us about. The second type of physical property we can call o-physical properties because they are the type of properties that would be required if we were to give a complete account of the intrinsic nature of paradigmatic physical objects. Physicalism can then be the less inclusive t-physicalism (that cannot account for causal efficacy), or the more inclusive o-physicalism.

For the sake of argument, we should grant that the uses of ‘physical’ in ‘physical theory’ and ‘paradigmatic physical object’ are cogent (see Stoljar 2001a, fn 9-10). It may be useful, to get a hold on the notion of o-physicalism, to think of ‘paradigmatic physical objects’ as ‘paradigmatic non-mental objects’, as discussed by Melnyk (1997), Crook and Gillett (2001) and Spurrett and Papineau (1999), such that physicalism at least is the claim that the physical, or non-mental, constitute the mental. Granting this means that we accept that o-physicalism is a proper type of physicalism. This allows us to, in a later sections, debate whether or not we should believe o-physicalism is actually true, while setting aside debate about whether o-physicalism really deserves to be labeled as a type of physicalism. Notice however that both kinds of debates presumably would

be driven, in the first instance, by the observation that we have a very poor understanding of what we accept into our ontology when we accept the existence of o-physical properties (especially if defined merely in terms of properties of ‘paradigmatic non-mental objects’). For example, someone who worries about whether o-physicalism is really a type of physicalism could reasonably ask what substantial difference to physicalist doctrine it would make to accept “o-dualist” properties rather than “o-physicalist” properties.

(This cuts a long story rather short. You could also arrive at o-physicalism by first noticing that t-physics characterises only extrinsic, relational properties, and then wondering what it is that is related in the ways t-physics says. The natural answer is that it is the intrinsic properties that are related in those ways. For our purposes it is the considerations about causal efficacy that are crucial.)

The mind–body debate

Stoljar very usefully discusses the mind-body problem in terms of this inconsistent tetrad:

- (1) if physicalism is true, then a priori physicalism is true
- (2) a priori physicalism is false
- (3) if physicalism is false, then epiphenomenalism is true
- (4) epiphenomenalism is false

Stoljar argues that if we use our new conceptions of the physical to disambiguate on the notion of physicalism in (1)-(4) (and if we decide to grant a number of controversial issues for the sake of argument), then there is no inconsistency. Here is how:

(1) if physicalism is true, then a priori physicalism is true

Physicalism is the doctrine that the mental supervenes with metaphysical necessity on the physical. Plausibly, the only consistent way to flesh out what supervenience comes to in this case is in terms of the a priori entailment of the mental truths by the physical truths (plus topic neutral truths). On this view, once we know all the physical truths, it will be inconceivable that there be non-conscious physical duplicates of us; hence the truths about the mental follow a priori from all the truths about the physical (Chalmers 1996, 2002; Jackson 1998).

The alternative is a posteriori identity relations between the physical and the mental, on a par with the a posteriori necessary relations found in fundamental laws of nature. But there is no explanatory rationale for making the identification in the case of the physical and the mental, in contrast to the case of fundamental laws where the relations are explanation-driven. Without an explanatory rationale, the identification just begs the question (for more, see Chalmers (2002) on ‘type-B materialism’).

(2) a priori physicalism is false

If a priori physicalism is true, then one could in principle know all the mental truths, including truths about qualia, on the basis of knowing the physical truths. But a couple of famous

arguments—the knowledge argument and the zombie argument (Jackson 1982; Chalmers 1996)—tell us that qualia are epistemically distinct from the physical: the core point is that there is no way the structural, functional properties of physics could constitute the intrinsic qualitative properties. Therefore, no amount of physical truths could suffice for the entailment of truths about qualia. So a priori physicalism is false.

But Stoljar invites us to consider:

(2-o) a priori o-physicalism is false

This says that it is false that qualia supervene on o-physical properties and that the o-physical truths entail the qualitative truths a priori. The considerations of epistemic distinctness do not give us reason to believe (2-o).

Mary in the black and white room may know all the t-physical truths and yet fail to know all the qualitative truths. But Mary may fail to know all the qualitative truths because she fails to know all the o-physical truths. Thus, for all the knowledge argument says, a priori o-physicalism may be true.

A core premise in the zombie argument is that I can conceive of a physical duplicate of me who does not have qualitative states. This seems true for a t-physical duplicate because we know a lot about the t-physical truths that may be relevant here. But it may be false for an o-physical duplicate:

since physical theory tells us nothing about categorical properties, we have very little understanding of their nature. Arguably, we can only conceive of things whose nature we have a proper understanding of (at least if we want to use conceivability as a guide to metaphysical possibility), so it is false that I can conceive of an o-physical twin of mine who is not conscious (Stoljar 2001b).

(3) if physicalism is false, then epiphenomenalism is true

If physicalism is false, then there are non-physical mental properties. If we believe in the causal closure of the physical, then non-physical properties can do no causing of the physical. But then these mental properties are epiphenomenal.

But here, Stoljar points out, it seems that the following is false:

(3-t) if t-physicalism is false then epiphenomenalism is true

If t-physicalism deals exclusively in dispositional properties, and we believe that dispositional properties are causally efficacious only via their categorical, o-physical bases, then we need to say that there are some physical properties, o-properties, that are causally efficacious. But, then, even though t-physicalism is false, epiphenomenalism may be false too: the causing may be done by o-physical events. On the other hand, if o-physicalism is false, there is nothing left to do the causal job of the mental, and epiphenomenalism will be true. (I return to this below).

(4) epiphenomenalism is false

Our evidence of qualia stems from causally based cognitive systems of memory, introspection and perception. Epiphenomenal properties could make no causal deliverances to such systems. But we do have evidence of qualia, so qualia must be causally efficacious properties, so epiphenomenalism must be false.

If we put all this together, with particular focus on the considerations about (2) and (3), we get the following consistent tetrad:

- (1) if physicalism is true, then a priori physicalism is true
- (2-t) a priori t-physicalism is false
- (3-o) if o-physicalism is false, then epiphenomenalism is true
- (4) epiphenomenalism is false

This is the elegant strategy for resolving the problem that Stoljar suggests. The disambiguation between o- and t-physicalism allows us to accept plausible versions of all four propositions. We now examine in more depth the reasons for believing that o-physicalism is true.

3. An inference to the best explanation?

None of the considerations that count against t-physicalism count against o-physicalism. O-physicalism can allow us to hold on to versions of (1)-(4), and we do not have to reject the compelling motivations behind the four original propositions. It appears no other candidate position comes anywhere near such stunning success.

Stoljar's primary aim is to show that o-physicalism is a real contender, in the sense that nothing currently tells us it could be false and that it make versions of the four propositions consistent. It is not his aim to make a positive case that o-physicalism is true and to explain the phenomenal, as demonstrated by the fact that his resolution to the problem as set out in (1), (2-t), 3-0) and (4), isn't formulated in terms of claims like:

(1-o) if physicalism is true, then a-priori o-physicalism is true,

or

(2*-o) a priori o-physicalism is true.

So, strictly speaking, we cannot criticise Stoljar for not providing positive grounds for believing o-physicalism. However, the next natural step, after having noted the o-physicalist resolution to the problem is surely to consider whether there could be good reasons to actually believe that o-physicalism is true. If we do not take address this further issue, then the acknowledgment of o-physicalism as resolution to the problem seems idle. So it is to this issue we now turn.

Given that we have no direct science of o-physical properties, it seems that, as Chalmers notes (2001, §11), the only way we could argue for a position like o-physicalism is by inference to the best explanation. And in fact, if we put everything together we could easily be led to infer that o-

physicalism would provide the best explanation of the occurrence of phenomenal experience: the metaphysical doctrine we ought to accept, when compared to its relevant rivals. Without o-physicalism we must reject one or more of the motivations for (1)-(4). Surely any explanation that allows us to keep them is better than an explanation that forces us to reject some of them. We can thus see that t-physicalism certainly doesn't work as well as o-physicalism; and that if o-physicalism is false then we get some kind of dualist epiphenomenalism that integrates very poorly with what else we know and believe.

Accordingly, it is legitimate to focus on o-physicalism's promise of explanatory prowess. O-physicalism is going to be the best explanation if we have reason to believe that it could put us in a better position with respect to the explanatory gap than can positions such as t-physicalism, neutral monism, and various forms of dualism. It can only be expected to do this job if there are no a priori obstacles to inferring that o-physicalism could be the *best* explanation; and if there are no reasons to doubt that o-physicalism could in fact *explain* the constitution of qualitative properties (it might be the best of a class of too poor explanations). I will raise concerns that suggest that the case for o-physicalism is not strong enough to do either job (thus, as I have said, I will not question the general metaphysical framework within which the problem and Stoljar's solution are phrased).²

I foresee the objection that we cannot evaluate the explanatory quality of o-physicalism because it is a philosophical doctrine, not an empirical theory. Response: (i) The philosophical doctrine of t-physicalism is not an empirical theory either, but we can nevertheless evaluate its explanatory quality; otherwise we would not believe there is an explanatory gap. (ii) We can evaluate o-

physicalism's qualities as an explanation sketch, or a conception of an explanation, against its relevant rivals, that is, other philosophical doctrines that similarly work as explanation sketches, or conceptions of explanations of phenomenality. (iii) We will not hold it against o-physicalism that it cannot offer structural and functional explanations in the style of t-physics, as that would beg the question; it would also be irrelevant, since we strongly suspect that such explanations are insufficient when it comes to phenomenality. However, we have to tread carefully. A core premise in the argument against the explanatory capacity of t-physicalism is the conceivability of a t-physical duplicate of us who is a zombie. This kind of conceivability arguably presupposes a proper understanding of the nature of t-physical properties (Stoljar 2001b). A directly analogous premise, about an o-physical zombie duplicate, is therefore not available in a critique of o-physicalism since we have as yet no understanding of the nature of o-physical properties. On the other hand, we do have some limited conception of o-physicalism: its explanations cannot be wholly structural and functional, and its characteristic properties are intrinsic and causally efficacious. We can therefore legitimately ask whether these attributes detract or add to our reasons for actually believing that o-physicalism is better than its rivals. Thus, consider the so-called grain problem (discussed by Stoljar 2001a: 275–277) that neutral monism-style positions like o-physicalism are inadequate because assemblies of intrinsic particles only seem able to constitute grainy phenomenal states, not the phenomenal states we associate with, say, the experience a smooth expanse of red. This is well recognised as a problem for the explanatory capacity of such positions, even though it could not be based on understanding of the nature of neutral properties or o-physical properties. However, this objection does involve conceiving of something like a partial o-zombie: a duplicate in neutral (or o-) properties who have grainy, not smooth, experiences.

4. O-physicalism is not the best explanation.

Consider the rival doctrine that we might call o-dualism that is much the same as o-physicalism, except its categorical properties, or some of them, are irreducibly mental. The case we have built up for o-physicalism is not strong enough to make us infer that it is better—holds more explanatory promise—than o-dualism.

First, the physicalist has to posit *physical* o-properties, if physicalism is going to be true. Likewise, the o-dualist has to posit some *mental* o-properties, if dualism is going to be true.

Second, o-dualism and o-physicalism are equally *simple* and *informative*. They are both content poor accounts of the qualitative, and they are therefore not very informative at all, and their simplicity is achieved through content poverty, not in spite of content richness.

Third, o-physical properties *integrate* well with the t-physical because they are needed to complete the physicalist's picture of the causal goings-on in the world. But the o-mental properties are also needed to complete the dualist's picture of causality (at least the kind of dualist who is impressed by the causal inefficacy of dispositional properties).

In another sense both o-physical and o-mental properties integrate very poorly with t-physical properties. We can have two t-physically similar worlds with entirely dissimilar sets of o-physical

properties.³ And it is the same with two t-physically similar worlds and different sets of o-mental (plus o-physical) properties.

Of course, o-dualist properties are weird and wonderful properties, and weird properties typically contribute to worse explanations than familiar properties. However, this doesn't mean that o-dualism is worse off than o-physicalism. First, o-physical properties are weird too. We have absolutely no understanding or science of them, nor have we any acquaintance with them. So it is premature to judge them less weird than o-dualist properties. Second, though o-dualist properties are weird, we are at least intimately acquainted with categorical, intrinsic mental properties that might well be o-dualist in nature. As Stoljar notes, (2001a: 273), we derive our very concept of categorical properties from our acquaintance with a type of intrinsic mental property, namely qualia.

Fourth, for an explanation to be better than its rivals, and not just the best out of a class of poor explanations, it must provide some sort of explanatory *mechanism* whereby the explanandum events occur. O-dualism doesn't have to provide a mechanism whereby the qualitative is constituted, since o-mental properties already are qualitative, though on the other hand we need a mechanism that can explain how these properties hook on to neurophysiological properties. This makes o-dualism a pretty poor explainer, I think. But neither are we given any idea about what the mechanism is in the case of o-physical properties and their constitution of the qualitative. All we know is that the mechanism is neither structural nor functional. This does not suffice to decide that o-physicalism provides better mechanisms than o-dualism (more on this later). Of course, we are told something,

namely that o-physical properties supply the causal efficacy to dispositional properties, but the o-dualist can say much the same about her o-mental properties.

The upshot is that nothing compels us to choose o-physicalism over o-dualism. For all I have said so far, it might in fact be capable of better explanations of the qualitative, but we have no reason to believe that now.

But there is an objection to this result. If o-dualism is true, then o-physicalism is false; and if o-physicalism is false, then epiphenomenalism is true ((3-o) above). But epiphenomenalism is false, so o-dualism is false. O-physicalism is after all better.

This objection depends on our reasons for believing (3-o), that is, our belief in causal closure of the physical. Our belief in the causal closure of the physical stems from the past explanatory success of physics. Those explanations are couched exclusively in t-physical terms—we do not have o-physical explanations yet. So, strictly speaking, and in the absence of any prior belief in the metaphysical doctrine of physicalism, our warrant for belief in causal closure does not reach any further than the t-physical. We might go on to say that, as physicalists, we believe that the causally efficacious properties are also physical. But if we in this manner presuppose physicalism we cannot argue in favour of physicalism.⁴ So considerations of causal closure does not suggest the truth of:

(3-d) If o-dualism is true, then epiphenomenalism is true.

What we should believe, if we believe in causal efficacy at all, is the fairly weak:

(3-?) if o-?-ism is false, then epiphenomenalism is true.

This proposition says that, if epiphenomenalism is false, then there are some categorical properties that are causally efficacious. But it is silent on the ontological status of those categorical properties. So, when we adjust the tetrad so it consists of (1), (2-t), (3-?) and (4) we have no reason to pick o-physicalism over o-dualism (or other positions).

Some might worry that this discussion of the causal closure of the physical is too wide-ranging.

Doesn't it imply that the physical is not causally closed one way or the other: if the t-physical needs o-physicals to be causal at all, then the relevant arena for discussing causal closure was always the o-physical? That might be true, but, again, if we focus on the actual evidence for the causal closure principle we see that it is silent on the issue of o-physicals. We have found that all our successful causal explanations of physical events are t-physical (no matter how deep down in the hierarchy of dispositions and their non-categorical realisers t-physics has probed). From the perspective of the evidence for the closure principle we must say that ontological conclusions about the categorical o-level are non sequiturs. I take this to be an important upshot of the discussion so far. This does not mean that we should cease to believe in the causal closure of the t-physical, that is, believe that there are no gaps in the causal chains described by t-physics.

A strengthened objection is that belief in causal closure, in *conjunction* with belief in the causal inefficacy of dispositions, rationally compels belief in the causal closure of the o-physical. The believer in causal closure of the t-physical can respond to this by adopting a version of t-physics where concepts of dispositional properties refer to their categorical grounds *inter alia*. On this version of t-physics, the classes of t-physical and o-physical properties are not distinct, and is of no use in the quest for o-physicalism—the classes need to be distinct if we are going to believe that o-physical properties can explain more than t-physicals (Stoljar 2001a: 260). The believer in causal closure can then, reasonably it seems to me, get causal efficacy, keep believing that there are no t-physical causal chain gaps, and yet be agnostic about the metaphysical status of o-properties. Epistemically reasonable belief in causal closure is therefore independent of considerations that lead to belief in o-physicalism.

5. Ignorance and o-physicalism.

An analogy developed in some detail by Stoljar is intended to throw positive light on the dialectic leading up to acknowledgement of o-physicalism (2001a: 269–270; 2001b: 401). We are told to imagine a mosaic constituted by two basic shapes: triangles (read: t-physical properties) and pieces of pie (read: o-physical properties). We have only two shape-detecting systems: one for detecting triangles and one for detecting circles (read: qualitative properties). We come to believe that all other shapes supervene on triangles. But then the problem arises that there is no place for circles in a world of triangles. There is in other words an explanatory gap between the story about triangles and the story about circles. In response one can adopt parallels to the traditional positions in

philosophy of mind (saying that we have no propositional knowledge about circles, that circles are a posteriori identical to triangles, that there are contingent laws connecting triangles and circles etc.).

Those parallel positions all make the same mistake. They assume that the only viable supervenience thesis is the supervenience of all shapes on triangles because they mistakenly think that the triangle detector tells us everything. But it is selective (just as t-physics is selective). If we were able to detect pieces of pie, which are not themselves circles, then it becomes reasonable to hold that all shapes supervene on triangles and pieces of pie.

By analogy, if we were to learn that in addition to t-physical properties there are o-physical properties, which are not themselves qualitative properties, then it would be reasonable to hold that qualitative properties supervene on t-physical and o-physical properties.

However, the analogy cannot be used to support belief in o-physicalism since there are relevant dissimilarities between the shape case and the physicalism case.

The analogy trades on the fact that surely we can agree that we would be entitled to believe that triangles and *pieces of pie* constitute circles. But we are so entitled *because* we can readily see that the explanatory gap concerning shapes would be bridged by adding pieces of pie to the triangles. It is not the other way around: that we can see that the explanatory gap is bridged *because* we are entitled to believe that circles are constituted by triangles and pieces of pie. If we were somehow to learn that in addition to the triangles there were straight lines, not pieces of pie, we would not

suddenly be entitled to believe that triangles together with straight lines constitute circles. The reason is that straight lines would not help bridge the explanatory gap. The mere fact that there exists a type of property that we had not hitherto detected doesn't do anything to make it plausible that those properties are relevant for constitution.⁵

Now we can see that the analogy doesn't help us to appreciate o-physicalism. The only thing that the o-physical story tells us is that there are categorical properties that are causally efficacious. But we are given no reason to suppose that such properties could bridge the explanatory gap. In the absence of such reason we are not entitled to believe that categorical properties go into the constitution of qualitative properties.

Of course, we are told something about o-physical properties, viz. that they are intrinsic and lend causal efficacy to dispositional properties. But this in itself does not suggest that the explanatory gap can be closed. In terms of the analogy, it is comparable to being told only that the new properties are geometrical, not that they are pieces of pie—many geometrical properties are irrelevant for the constitution of circles. Likewise, there are many different kinds of intrinsic property (being 6 ft tall may be one of them), and most of them will play no role in the constitution of qualitative properties. Further, though we do need to explain how epiphenomenalism can be false, the addition of mere causal efficacy is not at all tantamount to the constitution of qualitative properties.

Part of the point here is that o-physicalism might be true because there is something we are ignorant about when we try to conceive various scenarios (see Stoljar 2001b: 405–8). But if the point *just* concerns ignorance, then o-physicalism gets no support: we may be ignorant of so many types of things.

The conclusion so far is that it is doubtful whether o-physicalism has *better* explanatory promise than its relevant rivals, such as o-dualism. Whether o-physicalism is better in this sense or not, we must now address the equally important question whether there is good reason to believe it can furnish us with *explanations* of the phenomenal at all.

6. The constituting role of o-physical properties.

Chalmers and others have objected to materialist versions of the Russellian position that they look like panpsychism (Chalmers 1996: 154–5). Panpsychism, in this context, is the weird idea that since all categorical properties are qualia, everything instantiates qualitative properties. What motivates the claim that *all* categorical properties are qualia? The argument is that we get the concept of a categorical property from our concepts of qualia, so if all things have categorical properties, then all things instantiate qualia. Stoljar rightly says that this does not follow since some, but not all, categorical properties may be qualitative and we may derive our concepts of categorical properties from the ones that are qualitative (2001a: 273).

Let us set aside the question of panpsychism and ask instead what use Stoljar's response is to the doctrine he wishes to defend. If it is going to be physicalism, then the categorical properties that

form the supervenience base need to be some of the physical ones. Truths about those intrinsic categorical properties need to entail truths about other intrinsic categorical properties, the qualitative ones we got our concept of categoricalness from.

But is this feasible? Either all categorical properties are primitive or they are not. If they are all primitive, then one kind of categorical property cannot constitute another. In that case there is no supervenience of qualitative categorical properties on o-physical categorical properties. If not primitive, then we need some kind of idea about what kind of constitutive relation holds between the categorical o-physical properties and the categorical qualitative properties. The kind of constitution must be different from functionalist paradigms of constitution, otherwise there would be no explanatory gap. But what could this be?

O-physicalism leaves us with a picture of constitution where massive agglomerations of intrinsic o-physical properties in certain combinations give rise to categorical qualitative properties. If constitution is possible in this mereological fashion, and if it is not capturable in dispositional (that is, t-physical) terms, then I think the most constructive thing that can be said is that qualitative properties *emerge* from the o-physical properties. That is, they are genuinely novel properties. But if that is right then we ought to be discussing o-emergentism instead of a priori o-physicalism.⁶ If some versions of emergentism are compatible with materialism, then proposition (1) would come out false since emergence of novel properties is not a priori. Further, if novel emergent properties have novel causal powers, then the physical is not causally closed; this erodes the motivation for (3-o).

As long as there is no feasible model of what a priori o-physical *constitution* of categorical qualitative properties could be, we have no particular reason to believe that o-physicalism could in fact furnish us with explanations—whether good or poor—of the phenomenal.

7. What current neuroscience makes us believe about the right level of explanation.

Here is a further cause for concern about whether o-physicalism can furnish us with explanations at all.⁷ Recent neuroscientific studies (e.g., Tong et al 1998) show that the neural correlate of conscious experience of faces includes activity in the fusiform ‘face area’ (FFA), whereas conscious experience of places includes activity in the parahippocampal ‘place area’ (PPA). It is reasonable to believe that something about the activity in FFA and PPA, and their interconnections to other areas of the brain, constitute the conscious experience of faces and of places, respectively. If o-physicalism is true, then this systematic difference in conscious experience is constituted by differences in categorical o-physical properties. But it would be amazing if the intrinsic properties of the tiniest subatomic particles of the molecules constituting the cells in FFA and PPA were relevant for the explanation of such phenomenal differences in the experience of faces and places. In my understanding, current physics and neuroscience give us no reason to believe that there is the requisite systematic difference, at this miniscule level, between the ventral and dorsal pathways, and they give us every reason to believe that there are relevant differences at higher, functional levels. Differences relevant to such fine-grained conscious states just seem to fade away as we go down the levels of physical properties. But if this is right, then we have reason to doubt that o-physicalism

can help explaining qualitative properties because what we need to explain is, for example, why one kind of physical activity is associated with one kind of conscious experiences and not another.

Perhaps this objection begs the question in favour of current science (which is based on t-physics), but it seems to me to rely on the same kind of evidence as our belief in the causal closure of the physical. It is also somewhat unfair to Stoljar who is neutral on the issue about to what extent functional properties need to be involved, in addition to o-physicals, in the final account of consciousness (2001a, n19). But my point here is that we have reason to believe that o-physicals have no role to play.

8. Some speculations about dispositions and consciousness.

Though we have seen that o-physicalism is neither best, nor furnishes us with explanations of how the phenomenal is constituted, the original case for accepting a level of intrinsic o-properties is not completely devoid of good-making qualities, especially concerning mental causation and epiphenomenalism. So, in spite of the pessimism about the prospects for a priori o-physicalism, we might still wonder what role categorical properties could have for our understanding of what it is like to be in qualitative states. If they do have an explanatory role, then what is it?

If categorical properties have a role to play then it is as the categorical bases of dispositions. So it would make sense to be a dispositionalist about consciousness. From this perspective we should say that when a subject is in a qualitative state some of the subject's dispositions are manifested. The

truth in the discussion about o-physicalism is that there cannot be any qualitative states without suitable categorical properties to ground these dispositions.

However, I do think the issue about categorical properties connects interestingly with the two most prominent answers to the knowledge argument: the perspectivalist and ability responses (see, e.g., Perry 2001, Lewis 1990).

The perspectivalist response says that when Mary is released she doesn't get new knowledge, she obtains an introspective perspective on what she already knew (Lycan 2002); or she acquires new subjective concepts of phenomenal states (Loar 1997; Block 2002). Since Mary gets no new knowledge of states of affairs there is no threat to physicalism.

This response may not be viable because Mary arguably does get new knowledge in addition to her new perspective: she learns that *this* thing she is introspecting *is the same* as a particular qualitative property. For example, she not only applies her newly acquired subjective concept to think '*that* [attention to a mental image] is red', she also learns that it is the same as the qualitative property of red-experiences (see, e.g., Chalmers 2002).

Whatever the status of this response, it is natural to explain Mary's ability to adopt an introspective perspective on her red-experiences, or her acquiring a subjective concept, in terms of dispositions. On leaving the room and seeing red for the first time she is caused to have dispositions for red-experiences. Something in the interaction with red stimuli sets up the necessary categorical bases.

When there is something it is like for her to introspect her red-experiences, or when she applies her subjective concept, these dispositions manifest themselves.

The ability response says that when Mary leaves the black and white room she doesn't acquire any new propositional knowledge, she simply acquires some new abilities: for example, the ability to imagine what it is like to experience red.

This response may also not be viable: though it is correct that Mary acquires new abilities she also gets new propositional knowledge, for example that this is what it is like for other people to experience red (Jackson 1982, 1986, Braddon-Mitchell and Jackson 1996).

Whatever the status of this response, it is natural to explain her acquisition of those abilities in dispositional terms: when Mary leaves the room and sees red for the first time she acquires new dispositions for having red-experiences. These dispositions are multi-track in the sense that different stimuli can make them manifest. Her ability to imagine red is then partly explained by the manifestation of these dispositions, triggered by something else than the normal external stimuli. Something in the interaction with red stimuli caused there to be the appropriate categorical bases for these dispositions.

We can then, after all, give the categorical bases of the dispositions characterised by t-physics a role to play in the full story about consciousness. They are needed to explain certain abilities and perspectives or concepts, but arguably not Mary's new knowledge. This is no surprise, really,

because we only know one thing about these properties: that they give us causal efficacy. Then the only explanatory roles there seem to be for these o-physical properties are such that, even if we did have the requisite o-physical explanations, there would still be an explanatory gap between the physical, broadly conceived, and the qualitative. We would still not have an explanation of what it is like. Neither t-physicalism nor o-physicalism can give us an explanation of what it is like to be subject to the causal forces unleashed when one's dispositions manifest.

9. Concluding remarks.

O-physicalism is a position worth taking seriously because, as Stoljar clearly shows, it is not a target of the traditional arguments against t-physicalism. This provides some solace for those who are already physicalists, but who need reassurance that their belief in physicalism is not inconsistent with other things they believe. However, though o-physicalism is initially promising, it is not a better candidate for resolving the problem of consciousness than a position that posits irreducibly mental, categorical properties, instead of only physical categorical properties. In addition, there is reason to think it doesn't hold much promise as an explanation of the phenomenal at all. So there is little to find for those of us who are agnostic about physicalism, but would like to have independent reason to believe it.

All the arguments against o-physicalism are based on the intuition that rational accept of such a doctrine depends on its promise of explanatory prowess. There seems to be no explanation-independent parameters that can take us beyond mere recognition of the possibility of o-physicalism and towards rational belief in it. This result is not arrived at simply by the observation that we

cannot make empirical experiments directly on the workings of o-physicals, rather it is arrived at by considerations about whether these hypothesized, theoretical entities could do the job required, and do it better, than other entities.

Though there are serious doubts about o-physicalism, there is an important insight to be had. Qualitative properties need to be causally efficacious, and the best way to get that is by thinking of them as dispositional properties with categorical bases. The problem of consciousness then depends on understanding how the manifestation of such grounded *dispositional* properties can constitute qualitative properties.⁸

References:

- Block, N. 2002. Consciousness. In *Encyclopedia of Cognitive Science*. Nature Publishing Group; Macmillan Publishers Ltd.
- Braddon-Mitchell, D. and Jackson, F. 1996. *The Philosophy of Mind and Cognition*. Oxford: Blackwell.
- Chalmers, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- Chalmers, D. 2002. Consciousness and its place in nature. To appear in S. Stich and T. Warfield (eds.). *Blackwell Guide to Philosophy of Mind*. Oxford: Blackwell.
- Chalmers, D. 2002. Insentience, indexicality, and intensions. *Philosophy and Phenomenological Research*.
- Crook, S and Gillett, C. 2001. Why physics alone cannot define the 'physical'. *Canadian Journal of Philosophy* 31: 333–360.

- Feigl, H. 1967. *The 'Mental' and the 'Physical'*. Minneapolis: University of Minnesota Press.
- Foster, J. 1991. *The Immaterial Self*. London: Routledge.
- Jackson 1982. Epiphenomenal qualia. *Philosophical Quarterly* 32: 127–136.
- . 1986. What Mary didn't know. *Journal of Philosophy* 83: 291–295.
- Kim, J. 1998. *Mind in a Physical World*. Cambridge, Mass.: MIT Press.
- Levine, J. 1983. Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64: 354-61.
- . 2001. *Purple Haze*. Oxford: Oxford University Press.
- Lewis, D. 1990. What experience teaches. In W. Lycan, (ed.) *Mind and Cognition*. Blackwell.
- Loar, B. 1997. Phenomenal States. In Block, Flanagan and Guzeldere, *The Nature of Consciousness: Philosophical Debates*, Cambridge: MIT Press. Pp. 597–616.
- Lockwood, M. 1989. *Mind, Brain, and the Quantum*. Oxford: Oxford University Press.
- Lycan, B. 2002. Perspectival Representation and the Knowledge Argument. In Q. Smith and A. Jokic (eds.), *Consciousness: New Philosophical Perspectives*. Oxford: Oxford University Press.
- Maxwell, G. 1978. Rigid designators and mind-brain identity. *Minnesota Studies in the Philosophy of Science* 9:365-403.
- Melnyk, A. 1997. How to keep the 'physical' in physicalism. *The Journal of Philosophy* 94(12): 622–637.
- Papineau, D. 2002. *Thinking About Consciousness*. Oxford: Oxford University Press.
- Perry, J. 2001. *Knowledge, Possibility, and Consciousness*. Cambridge, Mass.: MIT Press
- Russell, B. 1927. *The Analysis of Matter*. London: Kegan Paul.

- Spurrett, D. and Papineau, D 1999. A note on the completeness of 'physics'. *Analysis* 59 (1): 25–29.
- Stoljar, D. 2001a. Two conceptions of the physical. *Philosophy and Phenomenological Research* 62: 253–81. Also in Chalmers, D. (ed.) *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press, 2002
- . 2001b. The conceivability argument and two conceptions of the physical. *Nous* 35(s15): 393–413.
- Strawson, G. 2000. Realistic materialist monism. In S. Hameroff, A. Kaszniak, and D. Chalmers (eds.), *Toward a Science of Consciousness III*. Cambridge, Mass.: MIT Press.
- Tong, F., Nakayama, K., Thomas Vaughan, J., Kanwisher, N. 1998. Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21: 753-759.

¹ See Russell 1927. Similar, but less overtly physicalist, positions have been discussed by Feigl 1967, Maxwell 1978, Lockwood 1989 and Strawson 2000.

² Some of these concerns concur with those of Chalmers 1996: 157, 307; and Chalmers forthcoming, though I arrive at them by considering the notion of explanation. They are also aligned with Levine's brief remarks in his 2001: 25-26, 177. See also Foster 1991, Ch. 4.4.

³ This is an aspect of what Jackson calls Kantian physicalism, Jackson 1998: 23.

⁴ Papineau (2002, Appendix) offers a different justification for the belief in causal closure, but this justification is even more explicitly couched in terms of t-physics, so the point about the scope of warrant remains.

⁵ But what is the role of the straight lines in the new version of the story? Are they just idle ontological danglers? They might just be danglers, but they could also constitute the triangles, just as the categorical o-properties might constitute the dispositional t-properties. But in the first case the problem of the circle would persist, just as in the latter the problem of consciousness would persist.

⁶ Foster (1991, p. 128) is also suspicious that a Russellian position is really committed to emergence.

⁷ This type of objection is also found in Foster 1991, p. 127-9.

⁸ Thanks to the audiences at the 4 European Conference for Analytical Philosophy, June 2001 in Lund, and at a seminar at the department of philosophy, University of Melbourne. And thanks to Daniel Stoljar for many discussions.