

Do Phenomenal Concepts Misrepresent?

Byrne, Darragh

DOI:

[10.1080/09515089.2015.1108398](https://doi.org/10.1080/09515089.2015.1108398)

License:

Other (please specify with Rights Statement)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Byrne, D 2016, 'Do Phenomenal Concepts Misrepresent?', *Philosophical Psychology*, vol. 29, no. 5, pp. 669-678. <https://doi.org/10.1080/09515089.2015.1108398>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

This is an Accepted Manuscript of an article published by Taylor & Francis in *Philosophical Psychology* on 22/01/2016, available online: <http://www.tandfonline.com/10.1080/09515089.2015.1108398>

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Do Phenomenal Concepts Misrepresent?

Darragh Byrne, University of Birmingham, UK

Abstract:

Many contemporary physicalists concede to dualists that conscious subjects have distinctive “phenomenal concepts” of the phenomenal qualities of their experiences. Indeed, they contend that idiosyncratic characteristics of these concepts facilitate responses to influential anti-physicalist arguments. Like some other critics of this approach, James Tartaglia maintains that phenomenal concepts express contents that conflict with physicalism, but as a physicalist, the moral he distinctively draws from this is that phenomenal concepts *misrepresent*. He contends further that the contemporary physicalists’ account cannot accommodate this feature, and that in consequence, physicalists should abandon phenomenal concepts and return to the “Identity Theory” championed by Place and Smart in the 1950s. I respond to Tartaglia by identifying lacunae in his interpretation of contemporary physicalism, and arguing that phenomenal concepts as conceived by the contemporary physicalists do not express contents that support either dualist or physicalist metaphysics: they are “metaphysically neutral”.

Most of us think that mental states and their properties are physical or functional, but very few would defend that metaphysical theory with an argument based on introspective evidence. Thoughts and sensations do not “present themselves” *as* physical. The correct metaphysical theory of consciousness is not “revealed” to conscious subjects through consciousness itself. A central component of a popular contemporary physicalist position is the idea that conscious subjects have distinctive “phenomenal concepts” that refer *directly* to

what it is like for them to undergo their experiences. According to James Tartaglia (2013)¹ this idea conflicts with the fact that experiences do not present themselves as physical. Tartaglia's reasons for holding this are natural, important, and prima facie compelling; however as I shall explain in this note, they are based on misunderstandings of the physicalists' conception of phenomenal concepts.

1. Tartaglia's Misrepresentation Thesis

As Tartaglia explains, the reason 1950s physicalists² were not threatened by the observation that sensations and other mental states do not introspectively appear to be physical is that the physicalists held that our concepts of all of these states are “topic neutral” and, in effect, *functional*: causal role concepts whose contents are “fixed extrinsically by the circumstances in which they typically occur rather than by anything intrinsic [to their referents]” (p. 818)³. In response, critics argued that at least for sensations, that conception of our introspective thinking is implausible. It seems that we can conceive of disembodied minds, zombies, qualia inversion etc., and that to do so we utilize concepts of “what it is like” for us to be conscious which are not functional concepts, and indeed which capture features that are intrinsic and essential to conscious experiences. Since the 1990s it has become fairly orthodox for physicalists to concede that point to the anti-physicalists — to concede that conscious subjects have distinctive “phenomenal concepts” (hitherto, ‘PC’s) of what it is like for them to undergo their experiences, and that these concepts do not have functional contents. These contemporary physicalists⁴ (whom I shall hitherto label the ‘PC theorists’) maintain that what is especially distinctive about PCs is that they refer *directly* to their (physical or functional) referents. Most concepts, they assume, refer by way of association with (accidental) features of their referents: e.g. the concept, WATER is associated with water’s liquidity, transparency, potability etc., and its referent is fixed as the substance that happens to exhibit those features

in the actual world.⁵ In contrast, say the PC theorists, the contents of PCs are (all but) exhausted by their referents. This relation of direct reference allows the PC theorists to concede to their opponents that by expressing what it is like to undergo experiences, PCs represent intrinsic and essential properties of those experiences. However they argue that direct reference also implies that (and explains why) PCs are not a priori connected to the “physical concepts” we use to think about physical/functional properties as such, and that this in turn allows us to explain away the epistemic data that feature in influential anti-physicalist arguments in a manner which is consistent with physicalism about PC’s referents. Thus, e.g. it explains why zombies, disembodied minds and qualia inversions are conceivable (though metaphysically impossible).

Why does Tartaglia think that introspection’s failure to reveal the physicality of experiences is problematic for these contemporary physicalists? The answer is that the physicalists’ account of this introspection is in terms of PCs, and according to Tartaglia, PCs not only fail to reveal that their referents are physical — worse, they represent them as *non-physical*: “phenomenal concepts are dualist concepts” (p. 820), “loaded with ontological significance: if you conceptualize conscious experience in this way, you will inevitably [...] represent it as having a nature that physicalism cannot accommodate” (p. 828). This idea is not altogether new. E.g. Nida-Rümelin (2007) proposes that possession of concepts of various kinds – including PCs – affords subjects with “grasps” of properties – where to “grasp” a property is to “understand what having that property essentially consists in” (2007, p. 307); and she defends property dualism by arguing that subjects who possess PCs and also appropriate “physical concepts” can know by dint of these grasps (and so a priori) that they are not co-extensive. In a related spirit, Goff (2011) defends an epistemic thesis slightly weaker than Nida-Rümelin’s, i.e. that PCs are “translucent”, which is to say that each one “reveals part (but not all) of the nature of its referent” (2011, p. 194); and he argues that the

physicalist position I described above is not consistent with this. But although Tartaglia would sympathize with these philosophers' contentions that PCs present their referents in a way we cannot reconcile with physicalism, the distinctive thing about his position is the idea that in doing so, they *misrepresent* those referents. According to his version of physicalism, there simply are no properties of the kind represented by PCs: "phenomenal qualities" are "mythological" (pp. 819, 820, 833).

Tartaglia's position is also distinctive in that it involves two arguments I have not encountered before. But before examining these, let us briefly take a closer look at Tartaglia's intriguing idea that phenomenal concepts misrepresent.

Prima facie, it is not altogether clear just how revisionary/eliminativist he intends his theory to be. His approach is deeply historicist, and he is interested in the historical evolution not only of philosophers' theoretical concepts of mind, but also that of our ordinary, "pre-theoretical" concepts. He suggests that our general concept of consciousness is a complex one which acquired a phenomenal component as a result of dualist thinking in the seventeenth century. But this component is dispensable, he thinks. Just as a primitive (person or community's) complex concept WHALE which (mis)represented whales as a kind of fish might evolve over time into one in which the fish component was replaced so that the concept came to represent whales as mammals, so our general concept of consciousness might evolve into one bereft of the phenomenal component: and what this remainder would amount to, he suggests, is a "topic neutral", functional concept of the kind championed by the 1950s physicalists.

Tartaglia's idea that our general concept of consciousness is in this way complex, with both phenomenal and functional components, is plausible, but it is worth asking where the phenomenal component came from. I presume the idea is that it is some kind of abstraction from individual PCs of individual phenomenal qualities⁷; but Tartaglia's account

of the latter is contentious. At times he seems to suggest that they too are anachronistic inheritances from dualist theory, and while that is unsurprising given his belief that they have dualist contents, it seems to beg the question against the PC-theorists' contention that their contents are exhausted by their referents. According to their advocates, PCs are distinctive in that they are not protean units of social cognitive currency. They are personal/individualistic — acquired not through immersion in a culture, but through episodes of introspective attention. If Tartaglia's view is that there are no phenomenal qualities, then given the way that PC theorists characterize phenomenal concepts, he ought perhaps to hold not that they misrepresent, but that there are none.

An alternative is to interpret Tartaglia as holding that the more specifically-focussed individual PCs which inform our general concept of consciousness are more complex than PC theorists suppose: as concepts that (mis)represent experiences as having phenomenal qualities, but which also represent further, functional properties of those experiences. Tartaglia writes of PCs referring to “experiences” (p. 3) and “conscious states” (p. 17), and as these are, plausibly, complex entities, this supports this interpretation. However the advocate of PCs is likely to complain that her idea was that PCs refer not to experiences or states per se, but merely to their phenomenal *properties*.⁸

At any rate, the contemporary physicalists that Tartaglia has in mind do not mean to construe PCs as misrepresentations. Their concession to philosophers critical of older forms of physicalism is that there is something it is (essentially) like to undergo an experience, and it is precisely that which they consider PCs to represent. Thus, as Tartaglia acknowledges, Loar (1990/97, p. 602) describes his approach as one which “take[s] the phenomenology at face value”, and other PC theorists would concur. If as these physicalists maintain, the content of a PC is exhausted by its referent, it is hard to see how a PC *could* misrepresent. It is even harder to understand how such a concept could have a specifically *dualist* content —

how it could somehow represent its referent as non-physical. I would suggest that the notions of physicality and non-physicality relevant to the metaphysics of consciousness are highly complex, theoretical philosophical ones which developed over centuries of intellectual evolution and whose precise significances are still very much foci of fraught debate.⁹

My view – which I think I share with all or most PC theorists – is that PCs do not misrepresent. But this is not because they represent their referents accurately as physical/functional – it is because they represent them neither as physical nor as non-physical. Even though they are not functional concepts, PCs are “topic neutral” as far as the metaphysics of their referents are concerned. I shall discuss Tartaglia’s two main arguments for the misrepresentation thesis in reverse order.

2. Tartaglia’s “Direct” Argument

Tartaglia regards his second argument for the misrepresentation thesis as the more “direct” (p. 821) of the two. It turns on the way in which advocates of PCs endorse the Cartesian claim which is also central to Kripke’s (1980) analysis, that phenomenal qualities are introspective *appearances*: there’s no appearance-reality distinction for phenomenal qualities. Although this is difficult to understand, it seems to fit well with the PC theorists’ conception of PCs as directly referring:

[We] do not conceptualize the underlying reality of a conscious experience in the way we conceptualize the underlying reality of a natural kind like gold indirectly on the basis of its golden appearance, because in the case of conscious experiences, the appearance *is* the reality being conceptualized. (p. 826)

Now, if phenomenal qualities are physical or functional, as physicalists maintain, it follows that certain physical/functional properties are introspective appearances. E.g. (following Tartaglia's use of a hoary old place-holder) if (the phenomenal quality of) pain is the stimulation of c-fibres, then since that phenomenal quality is no less and no more than an introspective appearance, it follows that c-fibre stimulation is an introspective appearance (pp. 828-9).

This consequence is puzzling, at least *prima facie*, and so Tartaglia spends some time investigating whether it might be underwritten by an illegitimate substitution into an intensional context. To see what he means, consider the following inference: *Hesperus is Phosphorus; Hesperus appears red; So, Phosphorus appears red*. If we suppose that Venus appears red under the (evening) mode of presentation associated with the name 'Hesperus', but not under the (morning) one associated with 'Phosphorus', then there is a clear sense in which this inference is invalid and has a false conclusion. However, Tartaglia urges, nothing analogous to this is going on in the phenomenal case we are considering, because the introspective appearances of phenomenal qualities as conceived by advocates of PCs are not contingent modes of presentation of them in a manner analogous to that in which the evening-time red appearance of Venus is a contingent mode of presentation of it.

Up to here I am in agreement with Tartaglia, and I concede that the idea that introspective appearances are physical properties is hard to understand. But I worry that his next step, in which he derives a striking metaphysical conclusion, is much too swift:

This conclusion is incompatible with physicalism, [...] because the concept of c-fiber stimulation, to which physicalists must grant sole ontological authority, has nothing to do with introspective appearance; a scientist would not have to mention introspection in order to explain what c-fiber stimulation is. (p. 829).¹⁰

How plausible is this? To explain what the property of c-fibre stimulation is, would a scientist have to mention introspective appearance? And if not, does that support Tartaglia's misrepresentation thesis? As I conceded above, the idea that introspective appearances are physical properties is *prima facie* puzzling. On the other hand I think we can dispense at least some of that *prima facie* puzzlement by reminding ourselves that the physicalist regards phenomenal qualities as physical *or functional*, and to me at least, the contention that an appearance is a functional property (even one with physical realizers) seems a good deal easier to countenance than the idea that it is a categorical physical property. I suggest that our intuitive conception of firing c-fibres is a conception of something rather categorical and concrete – cellular threads conducting electrical signals? – but of course 'firing c-fibres' was only ever meant to be a place-holder for a more sophisticated neurophysiological description, and the idea of a *functional* neurophysiological property is altogether more abstract. Appearances are, I suggest, *relations* of a certain kind between agents and the subject-matters of their perceptions (or here, introspections) and while it is very difficult intuitively to understand how vibrating cellular threads could amount to relations, the idea that functional properties (albeit with categorical physical realizers) could be perceptual relations is easier to countenance.

I have been arguing that Tartaglia's appropriation of the popular place-holder, 'firing c-fibres' threatens to distort the debate a little by distracting us from the likelihood that the best physicalist account of phenomenal properties might well turn out to be one according to which they are *functional* rather than categorical physical properties. To indemnify against this distortion for the remainder of this note I shall use the neutral label 'n' to denote the physical or functional neurophysiological property that, according to physicalists, will one-

day be identified with the phenomenal character of pain. So our question is: to explain what *n* is, would a neuroscientist need to mention introspective appearance?

I suggest that the answer we give to this depends on what we mean when we write of a scientist ‘explaining what *n* is’, but that no plausible interpretation of this supports Tartaglia’s contention that PCs misrepresent. Scientists’ explanations of things typically utilize scientific concepts – paradigmatically, the “physical concepts” we use to think about physical/functional properties as such, and which we encountered earlier as paradigmatic examples of concepts from which PCs are held by their advocates to be conceptually insulated. One way to address the question whether a scientific explanation of *n* would have to mention introspective appearance is to ask whether the concept, INTROSPECTIVE APPEARANCE is a “physical concept” of this sort.

If INTROSPECTIVE APPEARANCE is itself a *phenomenal* concept, or at least a concept with a phenomenal component or components (like the general concept of consciousness we considered in §1) then we can expect PC theorists to insist that it is not a physical concept, and I suspect Tartaglia would agree. On the other hand, I do not think it is entirely obvious that INTROSPECTIVE APPEARANCE is a PC. Perhaps it belongs to the theories of folk psychology and scientific psychology, and will one day be analysed in functional terms; perhaps it is a concept that someone with no experience of introspective appearances (like a version of Jackson’s (1982) Mary) could, in principle, acquire. And if INTROSPECTIVE APPEARANCE is a scientific, physical/functional concept, then physicalists can insist that Tartaglia is just wrong to assume that a scientific explanation of *n* would not mention introspective appearance.

On the other hand, the fact that we have more finely-focussed PCs of individual phenomenal qualities suggests that INTROSPECTIVE APPEARANCE is also a PC, formed in part by abstraction from the finer-grained concepts, like the general concept of consciousness I

discussed in §1. And even if it is not, Tartaglia could simply transpose his claim into one involving one of these finer-grained concepts. E.g., since WHAT IT IS LIKE TO BE IN PAIN is, presumably, not a physical concept, a scientific explanation of *n* cannot be expected to make use of *it*. For brevity's sake, I shall for the remainder of this section ignore this complication and assume that INTROSPECTIVE APPEARANCE itself is not a physical concept.

In this case we can agree with Tartaglia that a scientist's explanation of *n* would not mention introspective appearance. But even if *n* is an introspective appearance, a physicalist might accept this result with equanimity if she thinks that to give a scientific explanation of what something is, a scientist does not need to mention *all* of its features. To this Tartaglia would surely respond that what the problematic contention we are considering says is not just that being the introspective appearance of pain is *a* feature of the property of issue – it is that that it is its most salient feature, (in some sense) its *essence*. As we saw above, PC theorists endorse the Kripkean contention that the phenomenal quality of pain just is an introspective appearance, and e.g. Loar is happy to explain his theory of PCs' direct semantics in terms of their affording "direct grasp of [the phenomenal quality's] *essence*" (1990/97, p. 609. Italics added). Indeed, Tartaglia writes:

[Contemporary physicalists] cannot respond that the physical concept of c-fiber stimulation tells us both what c-fiber stimulation and the introspective appearance of c-fiber stimulation is, but simply fails to conceptualize it as the latter, because it is committed to the claim that introspective appearance is the *essence* of c-fiber stimulation. (2013, p. 831. Italics added.)

To respond to this we need to ask about the notion of *essence* that is being invoked here. As Tartaglia (p. 827) mentions, when Loar uses the word 'essence', it is to illustrate his

contention that PCs do not refer by being associated with accidental properties of their referents (as, e.g. I suggested in §1, theorists like Loar take WATER to refer by way of its association with water's liquidity, transparency, etc.). Loar's comment is (in Hill's term, also quoted by Tartaglia (p. 828)) a "purely negative" one about how PCs do *not* refer – it is not intended as a metaphysically heavyweight pronouncement concerning the fundamental nature of consciousness.¹¹

We could I suppose push a little against this deflationary interpretation of Loar's invocation of essence by insisting that his view cannot be *entirely* "negative". A first pass suggestion here would be that if on Loar's view, PCs do not exploit accidental properties of their referents (i.e. of phenomenal qualities), then perhaps the idea is that they exploit essential properties of those referents. But even if this is correct, I do not think the physicalist should concede that a scientific explanation of what something is must mention even all of its essential properties. Let's consider a stereotypical example of a mooted essential property. Origin essentialists contend that an animal's specific biological origins (the identity of its parents, or the zygote from which it developed) are essential properties of it – but that thesis is metaphysical, not scientific, and it does not entail that a scientific account of a given token animal has to identify its parents (or the parts of its parents from which it developed). Perhaps a scientific explanation of a given token animal ought to include information about the *kinds* of origins that animals of its kind can have (e.g. which properties its parents have to exhibit, whether they reproduce sexually or asexually, and so on) but that is not the same thing as identifying a token animal's essential properties in the metaphysical sense in focus here.

In any case, I think it would be better to interpret Loar as suggesting not that PCs exploit essential properties of phenomenal qualities, but that that they do not exploit *properties* of their referents at all – after all, he contends that they refer *directly* to those

referents – i.e. to phenomenal qualities. I suggest that his talk of PCs' relations to essences is not intended to suggest anything more metaphysically substantial than this.

As philosophers we can of course continue to do substantial metaphysics, e.g. to address questions about essences; and it would be interesting to bring recent developments in this area of metaphysics to bear on phenomenal qualities. A good example is the contention by Kit Fine (1994, 1995) and others that the familiar *modal* characterization of essential properties – as properties that an entity could not lack – is too course-grained to capture the notion we ought to be interested in. For an indication of what they have in mind, consider the set whose only member is the individual man, Socrates. It is plausible that having Socrates as its sole member constitutes the *essence* of this set (though indeed, it is also a property which the set could not lack). On the other hand, the man, Socrates could not lack the property of being the sole member of that set, but it does not seem right to say that being the sole member of that set constitutes Socrates's essence. Fine and his followers suggest that for the finer-grained, hyperintensional notion we need, we should look to the notion of “ground” or “metaphysical explanation”: even though they are necessarily co-incident, the set is *grounded* in the man, and not visa-versa: the existence or identity of the man *metaphysically explains* that of the set, and not vice-versa. In this vein we might wonder whether what PCs express correspond to the essences of phenomenal qualities in either of these senses: i.e. whether they conceive features of phenomenal qualities that those phenomenal qualities could not lack, or whether they somehow express essences in the Finean, finer-grained hyperintensional sense. These are interesting metaphysical and semantical issues which I cannot begin to settle here,¹² but however they turn out, it would still I think be a mistake to assume that the disclosure of these essences is a *scientific* enterprise. Just as, I argued above, origin essentialism does not entail that a given animal's parents must feature in a scientific characterization of it, the contention that phenomenal qualities have essences in either of the

senses just mentioned would not entail that a scientific characterization of them would have to invoke those essences.

Let me put this another way. I have been arguing that if the concept, INTROSPECTIVE APPEARANCE is not a “physical concept” – not the kind of concept that features in scientific explanations – then a scientific explanation of “what n is” should not be expected to mention that n is an introspective appearance, even if being an introspective appearance is essential to n. It does not follow from this that a *complete* characterization of n could fail to mention that it is an introspective appearance. In the final sentences of the section in which he presents the argument I’ve been considering, Tartaglia slips from talk of “scientific” explanation to talk of “complete” (p. 831) description/explanation, so perhaps this is what he has in mind. However at this point the physicalist should simply bite the bullet, and insist that her physicalism is the thesis that everything is physical, and not the stronger thesis that physics can explain everything on its own. From this perspective she can concede that a *complete* characterization of an introspective appearance would have to mention that it is an introspective appearance; however if as we have been assuming, INTROSPECTIVE APPEARANCE is not a “physical concept”, then that part of the characterization would be a detour into an area of enquiry that is not science: e.g. (as my own diversion into essentialism above might be taken to suggest) it would be a detour into metaphysics. It’s worth noting that some physicalists (e.g. Melynck 2003) characterize their position as the view that everything is explainable or describable by physics rather than as the claim that everything is physical, and this move would perhaps not be open to those physicalists. But to assume that this is not the position of the PC theorists I have been considering here is not *ad hoc*: one of the merits which those theorists explicitly claim for their theory is that it accounts for the fact that science cannot explain why experiences exhibit their phenomenal qualities (even though those qualities are physical).¹³

3. Tartaglia's Other Argument, and the PC-Theorists' Diagnostic Project

Tartaglia's other argument for his misrepresentation thesis emerges from reflection on the contemporary physicalists' view that a subject might unwittingly have two co-referring but conceptually independent concepts of a given phenomenal/physical property. Thus e.g. if we assume that pain is the neurophysiological (perhaps functional) property, *n*, we can think of these physicalists as contending that a person — let's call her Alice — might have a PC of how it is for her to be in pain, and also an *a priori* independent "physical concept" of *n*, which she uses when thinking about neurophysiology (as such), and that she might possess each of these concepts without realizing that they co-refer. Tartaglia concedes that a thinker might in certain cases entertain co-referring yet *a priori* independent concepts without realizing that they co-refer, but he insists that someone who did so in the phenomenal case in the way proposed by PC theorists would be "in error" (p. 824) in a way that would suggest that her PCs misrepresented their referents. "Concepts represent, so if there is no distinction in the properties of the represented referent to account for a distinction of *a priori* independent concepts, the distinction seems like nothing more than cognitive clutter" (p. 824).

I do not think we should diagnose misrepresentation or cognitive disorganization here. To fix ideas, we might compare the idea of distinct but co-referring concepts with a familiar one with which it obviously resonates – Frege's notion of distinct senses determining a common referent. Frege's point is not to diagnose a *defect* in our cognitive machinery (which might be corrected by somehow collapsing the distinction between sense and referent). Rather, it is to make sense of the idea that representations are perspectival – the idea that things can be thought about in different ways. Of course we can explain many familiar examples of this in terms of thinkers' sensitivities to distinct properties of referents – distinct concepts of Venus exploiting its celestial position in the evening versus its celestial position in the morning, for example – and to be sure, that particular kind of explanation is not

available here. But advocates of PCs offer an alternative explanation of cases like that of Alice: PCs do not exploit (accidental) *properties* of the phenomenal qualities that are their referents – they refer directly, exploiting the phenomenal qualities themselves.

Indeed, when we look more closely at the specific case of Alice, it is hard to find evidence of misrepresentation. Certainly, if what it's like to be in pain is *n*, then Alice lacks a true belief (i.e. the one she could express with the sentence, 'what it's like to be in pain is *n*') but this is *ignorance*, not error. Of course she might form the belief that what it's like to be in pain is *not n*, or even that it's not any physical or functional property. She would then have a false belief, which is I suppose a way to be "in error"; but there are two things we can note immediately in mitigation. First, either of these might be perfectly reasonable things for her to believe, given her overall evidence – given that as we have observed, conscious experiences do not *per se* provide evidence of their physical/functional natures. Second, and relatedly, even if it would be a kind of error for her to believe falsely that what it's like to be in pain is not *n*, it is far from clear why we should agree that this false belief would be engendered by a defect in the *concepts* involved, rather than, e.g. a deficit in her knowledge of neurophysiology or a mistaken assessment of the philosophical arguments for and against physicalism in the philosophy of mind.

By way of an analogy, Tartaglia discusses a variation of Kripke's (1979) 'Paderewski' example which is considered by Block (2007). Here, a subject associates the name 'Paderewski' with two distinct sets of properties of a single man, and so mistakenly believes himself to be speaking and thinking about two distinct men. Over time he forgets about the properties, and so even if his concepts of Paderewski were once connected, they come to be conceptually independent. Tartaglia regards this as another case of error – analogous to the pain/*n* case — but once again, it is not clear why we should agree with him. Certainly this latter subject lacks some true beliefs (e.g. one he could express with the

sentence ‘Paderewski is Paderewski’) and he has some false beliefs (e.g. ones he would express with, ‘Paderewski is not Paderewski’, ‘I know of two people called “Paderewski”’ etc.) but this all seems perfectly reasonable in the circumstances: his beliefs fit his evidence. And (most importantly) it does not seem plausible that either of his Paderewski concepts *misrepresents* its referent. If either one misrepresents, why not the other?

I have been arguing against Tartaglia’s contention that a subject like Alice would be “in error” in a sense that supports his thesis that PCs misrepresent, and in the course of my argument above I suggested that even if a given PC refers directly to a certain physical/functional property, a subject’s false belief that there are two distinct properties here might nonetheless fit her overall evidence. It’s important to emphasize that I do not mean here to suggest any concession to Tartaglia’s view that PCs present their referents as non-physical. To be sure, if they did so, that would be a way for a subject to have evidence that the properties are distinct; but my claim is once again rather that PCs are metaphysically *neutral*. As I argued in the previous section, their contents confer evidence neither for nor against the physicalist identifications at issue.

Towards the end of the section of his paper in which Tartaglia makes the argument for the misrepresentation thesis that I have been assessing in this section, he notes that certain PC theorists maintain that the distinctive characteristics of PCs help to explain *why physicalism is hard to believe*. One might take this as suggestive of the idea that PCs confer evidence against physicalism, and so as indicative of a discrepancy between the views of those PC theorists and the line I have taken here.

However, this would be an exaggeration. It’s true that some advocates of PCs invoke the concepts’ mooted special characteristics not only (as we saw in §1) to rebut anti-physicalist arguments, but also to prosecute what I have elsewhere (MS b) distinguished as a separate “diagnostic project” of explaining why physicalism about consciousness is difficult

to believe (even for philosophers who are convinced on theoretical grounds that it is true). Papineau (1993; 2002, Ch. 6; 2007) is the most prominent proponent of this, though he credits Loar (1990/97, p. 605) with a version of the same idea. His ambition is to explain why a peculiar feature of PCs makes it hard for subjects to that physicalism is true, but that the reasoning this involves exhibits a certain distinctive kind of fallacy. It would be an overstatement, I think, to describe this as a concession that the contents of PCs confer evidence against physicalism.

Papineau's (and Loar's) diagnostic proposal is built on an intriguing consequence of the particular accounts of PCs which they defend. This consequence is that there is a phenomenal dimension to the act of deploying a PC in thought – it is, phenomenally, *like* having an experience of the kind one is thinking about. Papineau suggests that PCs literally “quote” (or “use”) instances of phenomenal properties to refer to (i.e. “mention”) those properties. He considers a subject who entertains a proposition such as the one Alice might express with the words, ‘what pain is like is n’ – where we suppose that in doing so she deploys both a PC, WHAT PAIN IS LIKE, and the physical concept, N. His view is that the way in which the first of these conceptual deployments exhibits a phenomenal dimension while the second does not makes it hard for such a thinker to believe that she is thinking about a single property, because it leads her to commit a kind of use/mention fallacy. Concepts like N do not *use* a phenomenal quality to refer, so in contrast to PCs they seem to “‘leave out’ the experience at issue” (Papineau 2002, p. 170). But it's a fallacy: the fact that physical concepts do not use phenomenal qualities to refer to phenomenal qualities does not imply that they cannot refer to them. Notice that Papineau's diagnosis turns on a *negative* claim: physical concepts appear to “leave something out” because of a way in which they do *not* refer. The diagnosis is not the *positive* claim PCs somehow present their referents as non-physical.¹⁴

4. Conclusion

The idea that introspection delivers direct awareness of phenomenal qualities seems all too naturally to suggest that if those qualities are physical, then they ought to appear as such to introspectors. Tartaglia provides a valuable and instructive development of this idea, and his historical perspective on the matter makes his case especially interesting. However, as I hope to have established in this note, his arguments are informed by misunderstandings of the contemporary physicalists' conception of phenomenal concepts. If PC theorists take sufficient care to distinguish the metaphysical, epistemic and semantic components of their position, they can retain the entitlement to a contention that seems perfectly natural in any case: that physicalism is a doctrine of theoretical metaphysics that is neither confirmed nor disconfirmed by phenomenal experience itself.¹⁵

¹ All unqualified page references are to this paper.

² Tartaglia is thinking in particular of the "identity theorists" U.T. Place (1956) and J.J.C. Smart (1959).

³ I write "in effect" because this interpretation of the 1950s theory as a version of functionalism is somewhat charitable. E.g. the programme of providing "topic neutral" analyses of sensation terms which Smart (1959) proposed is not explicitly functionalist, though he later came to endorse a functionalist gloss. (See e.g. his (2012).)

⁴ Here Tartaglia and I are thinking of, e.g. Loar (1990/97), Hill (1997), Tye (1999) and Papineau (1993, 2002, 2007).

⁵ This is of course consistent with the orthodox Krippean insistence that these features do not constitute the meaning of the term 'water' or indeed, the content of the concept, WATER.

⁷ For more on the difficulties that arise here, see Papineau (2007, pp. 120-3).

⁸ If for the sake of argument we were to grant to Tartaglia his conception of complex concepts of this kind, his suggestion that they misrepresent their referents comes to look more like the contention that some of their components *have no* referents. And if that's right, his view is closer to eliminative materialism (a view sometimes attributed to Richard Rorty (1965) – the most historicist of the historicist philosophers from whom Tartaglia (p. 820) draws influence) than the kind of "reformed" physicalism his remarks suggest.

⁹ Given his enthusiasm for historicism in the philosophy of mind, Tartaglia strikes me as recklessly ahistoricist about the theoretical concepts, PHYSICAL PROPERTY and NON-PHYSICAL PROPERTY.

¹⁰ Tartaglia's mention early in this passage of the *concept* of c-fiber stimulation is I think a mistake, which we can charitably overlook. The contention that Tartaglia claims is incompatible with physicalism – that c-fiber

stimulation is an introspective appearance – is a contention not about a concept, but about a concept’s referent. A physicalist of the kind we are considering here would insist that the alleged fact that a scientist’s *concept* of c-fibre stimulation does not represent that property as involving introspective appearance does not imply that the property does not involve introspective appearance. (As I said, I shall assume that this is a mere mistake on Tartaglia’s part, but it is worrying that he repeats it on the next page: “[...] within a physicalist framework, there can be no difference between c-fiber stimulation and the introspective appearance of c-fiber stimulation. However, we already know there is a difference because the physical concept does not represent the property as an introspective appearance.” (p. 830))

¹¹ Goff (2011, pp. 198-9) makes a similar interpretative point about Loar’s remark, albeit in a different dialectical context (for more on which see §1 above).

¹² Presumably Tartaglia could not endorse the idea that PCs correspond to essences of phenomenal qualities in either of these senses, since his position is that there are no such qualities. He might, I suppose, maintain that they *purport* to do so.

¹³ In this section Tartaglia also touches on a version of physicalism he associates with Nagel and G. Strawson, according to which “our [scientific] understanding of physical nature [and of particular physical properties] is currently inadequate” (p. 827) because it cannot accommodate consciousness; so to do so, these concepts will need to be revised. He regards this as an alternative to the position of the PC theorists but rejects it on the somewhat telegraphic grounds that it “mystif[ies] physicalism with consciousness” rather than “demystify[ing] consciousness with physicalism”. The moral I have derived is not intended to be a version of this. First, the position I have represented is a version of the PC theorists’ view, not an alternative to it; second, as I understand it, the theory Tartaglia is alluding to anticipates that once our physical concepts have been suitably revised we shall be in a position to explain consciousness scientifically, whereas the theory I am representing says that to explain consciousness we need to do metaphysics (and semantics etc.) as well as science.

¹⁴ Tartaglia interprets this strand in these philosophers’ work as a contention that “dualist intuitions are hard wired” (p. 825), and he complains that this is something we should be reluctant to accept, since Cartesian dualism is a relatively recent innovation of modern Western philosophy. I agree with him that we should resist the idea that dualist intuitions are hard wired, but as I hope to have demonstrated above, this is not the diagnostic proposal we find in the work Papineau and others. (Note that I’ve been defending Papineau here for argument’s sake only. As it happens, I’m sceptical both about the efficacy of his diagnosis (see my (MS b)), and also about accuracy of the particular accounts of PCs on which it is based (see my MS a).)

¹⁵ For help with this piece I am especially grateful to James Tartaglia, with whom I have been discussing these issues for years. I am very grateful too to Naomi Thompson and an anonymous referee for *Philosophical Papers*.

References

-
- Block, N., 2006: Max Black's Objection to Mind-Body Identity, in T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford: Oxford University Press.
- Byrne, D., MS a: Phenomenal Senses.
- Byrne, D., MS b: Phenomenal Justification and the Intuition of Distinctness.
- Fine, K., 1994: Essence and Modality: The Second Philosophical Perspectives Lecture, *Philosophical Perspectives*, Vol. 8, Logic and Language, pp. 1-16.
- Fine, K., 1995: Senses of Essence, in Sinnott-Armstrong, W, D. Raffman & N. Asher, *Modality, Morality and Belief: Essays in Honour of Ruth Barcan Marcus*, Cambridge: Cambridge University Press.
- Goff, P., 2011: A Posteriori Physicalists Get Our Phenomenal Concepts Wrong, *Australasian Journal of Philosophy*, 89(2), pp. 191-209.
- Hill, C., 1997: Imaginability, Conceivability, Possibility and the Mind-Body Problem, *Philosophical Studies*, 87, pp. 61-85.
- Jackson, F., 1982: 'Epiphenomenal Qualia', *Philosophical Quarterly* 32: 127-136.
- Kripke, S., 1979: A Puzzle About Belief, in A. Margalit (ed.) *Meaning and Use*, Dordrecht and Boston: Reidel.
- Kripke, S., 1980: *Naming and Necessity*, Oxford: Blackwell.
- Loar, B., 1990/7: Phenomenal States (revised version), in N. Block, O. Flanagan & G. Güzeldere (eds.) *The Nature of Consciousness: Philosophical Debates*, Cambridge, Mass.: MIT Press.
- Melynk, A., 2003: *A Physicalist Manifesto: Thoroughly Modern Materialism*, Cambridge: Cambridge University Press.
- Nida-Rümelin, M, 2007: Grasping Phenomenal Properties, in T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford: Oxford University Press.
- Papineau, D., 1993: Physicalism, Consciousness and the Antipathetic Fallacy, *Australasian Journal of Philosophy*, 71.
- Papineau, D., 2002: *Thinking About Consciousness*, Oxford: Oxford University Press.
- Papineau, D., 2007: Phenomenal and Perceptual Concepts, in T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford: Oxford University Press.
- Place, U.T., 1956: Is Consciousness a Brain Process, *British Journal of Psychology* 47, pp. 44-50.

Rorty, R., 1965: Mind-Body Identity, Privacy, and Categories, *Review of Metaphysics* 19, pp. 24–54.

Tartaglia, J. 2013: Conceptualizing physical consciousness, *Philosophical Psychology*, 26:6, pp. 817-838.

Tye, M., 1999: Phenomenal Consciousness: The Explanatory Gap as a Cognitive Illusion, *Mind*, 108, pp. 705-25. Reprinted in his *Consciousness, Color and Qualia*, Cambridge, Mass.: MIT Press (2005).

Smart, J.J.C., 1959: Sensations and Brain Processes, *Philosophical Review*, 68, pp. 141-56.

Smart, J. J. C., 2012: The Mind/Brain Identity Theory, *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/win2012/entries/mind-identity/>>

Tartaglia, J., 2013: Conceptualizing Physical Consciousness, *Philosophical Psychology* 26:6, 817-838.