Mad Qualia

Umut Baysan

Abstract: This paper revisits some classic thought experiments in which experiences are detached from their characteristic causal roles, and explores what these thought experiments tell us about *qualia epiphenomenalism*, i.e. the view that qualia are epiphenomenal properties. It argues that qualia epiphenomenalism is true just in case it is (nomologically) possible for experiences of the same type to have entirely different causal powers. This is done with the help of new conceptual tools regarding the concept of an epiphenomenal property. One conclusion is that it is not obvious if qualia epiphenomenalism is false; and it is also not obvious what should make us believe that it is false---or for that matter, true. Connections between qualia epiphenomenalism, physicalism and nonphysicalist property dualism are further explored.

Keywords: absent qualia; inverted qualia; physicalism; properties; qualia epiphenomenalism; realization

I. Introduction

I want to revisit some classic thought experiments in which we are asked to imagine cases where experiences are detached from their characteristic causal roles. My aim is to show that we have so far failed to see what these thought experiments tell us about whether experiences have epiphenomenal properties; this I do with a new set of conceptual tools that I propose we should employ regarding the concept of an epiphenomenal property. For the reasons that I will elaborate on, I think we should pay special attention to the case of the madman and his mad pain. So, let me open with this case:

> [The madman is] a strange man who sometimes feels pain, *just as we do*, but whose pain differs greatly from ours in its *causes and effects*. … His [pain] is caused by moderate exercise on empty stomach, … turns his mind to mathematics, facilitating concentration on that but distracting him from anything else. Intense pain has no tendency whatsoever to cause him to groan or writhe, but does cause him to cross his legs and snap his fingers. He is not in the least motivated to prevent pain or get rid of it. (Lewis 1980: 216, emphasis added)

Lewis, in imagining mad pain, alongside Martian pain (which is identical to our pain in how it feels, what kinds of events it is caused by, and what kinds of events it causes, but is grossly different in its physical realization), aims to show that neither a simple functionalist theory nor a simple (psycho-neural) type-identity theory can be true. The former fails to accommodate mad pain, the latter Martian pain. The correct theory of mind-body relations should account for both mad pain and Martian pain being pains, Lewis argues. My aim in reopening the case of mad pain is not to assess functionalism, the type-identity theory, or Lewis's own proposal. I am interested in what the story of the madman tells about the causal role of the painfulness of pain.

By 'painfulness of pain', I mean a property of a mental state. The mental state in question is a state token (i.e. a particular event), not a state type. This property (painfulness) belongs to a group of properties that are properties of some mental states in virtue of which those mental states are experiences. When there is something that it is like to be in a mental state M, we say that M is a phenomenally conscious mental state (or simply, M is phenomenally conscious). When there isn't anything that it is like to be in M, we say that M is not phenomenally

conscious. Although nothing hinges on this assumption, for ease of articulation of certain claims, I will take it that all and only phenomenally conscious mental states are experiences. The properties that make mental states experiences are typically called 'qualia'; so by 'qualia', I will mean such properties. The bearers of qualia are experiences, which are particular events. Surely there are qualia tropes whenever phenomenally conscious mental state types are tokened; but in the terminology I employ here, qualia are themselves property types (and not tropes). My question here is what it takes for qualia to be epiphenomenal.

Let *qualia epiphenomenalism* (QE) be the view that qualia are epiphenomenal properties (i.e. properties without causal powers). It is important to clarify that QE is *not* the view that experiences cannot be causes. Let's call this latter view SQE ('S' for 'Strawman'). I will assume, at least for the sake of this paper, that SQE is *obviously* false. We say that the madman's pain differs from our pain in its causes and *effects*; we say that it *causes* the madman to cross his legs and snap his fingers. In so doing, we attribute causal powers to a pain experience. More generally, we are happy to accept that experiences can be causes (or parts of causes). So, it is worth emphasising that QE is not the view that experiences are epiphenomenal; it is the view that *the properties of experiences in virtue of which experiences are experiences are epiphenomenal*. There is a relevant distinction here between *token*-epiphenomenalism and *type*-epiphenomenalism (see McLaughlin 1989). Whereas token-epiphenomenalism holds that mental events (as particular instances of event types), are causally inefficacious, type-epiphenomenalism holds that even if such events may have causal powers, their falling under mental types have no role in this. I believe that SQE is a form of token-epiphenomenalism (applied to *experiences*), whereas QE is a form of type-epiphenomenalism (applied to *qualia*). This is in line with my suggestion above that qualia are property types.

QE may be false. But it is not *obviously* false. Moreover, not only is it not obvious that it is false, it is also not obvious what should make us believe that it is false---or for that matter, true. In what follows, I will explore this. This will show that even when philosophers associate (rightly, I shall argue) nonphysicalism about qualia with QE, they do it for the wrong reasons. The link between nonphysicalism and QE is subtler than many of us have thought. Showing this will also reveal surprising results regarding the cases of absent qualia and inverted qualia.

II. From Nonphysical to Epiphenomenal?

The question of QE famously comes up in Jackson's (1982) presentation of the knowledge argument against physicalism. I anticipate that anyone who is reading these pages will be familiar with the knowledge argument and the thought experiment about the captive neuroscientist Mary, so I will be very brief in my exposition. We are told that prior to her release from the black-and-white room, Mary knows all physical facts about colour perception; after her release, she learns something new about colour perception (namely what it is like to see something red); therefore, there are nonphysical facts about colour perception. From these, we are asked to conclude that qualia are nonphysical properties. After all, what was left out from Mary's complete physical knowledge of colour perception was her knowledge of colour qualia, something she acquires after her release. If colour qualia were physical, her knowledge of all physical facts about colour perception would have covered them. Or so argues Jackson.

For the sake of exploring its consequences, let's accept that the knowledge argument is sound and that qualia are nonphysical properties. Curiously, Jackson associates this conclusion with the view that qualia are epiphenomenal. Having argued that qualia are nonphysical properties, he feels the burden to show that it is not so implausible that qualia are epiphenomenal

properties. What is not immediately clear, to me anyway, is why qualia have to be epiphenomenal if they are nonphysical properties.

We see an explanation of the link from qualia nonphysicalism to QE in Lewis's (1988) discussion of the knowledge argument. Lewis argues that 'if something nonphysical sometimes makes a difference to the motions of physical particles, then physics as we know it is wrong … To believe in the phenomenal aspect of the world, but deny that it is epiphenomenal, is to bet against the truth of physics' (ibid.: 95). So, if 'phenomenal aspects' (qualia) are nonphysical properties, they had better be epiphenomenal, argues Lewis.[1]

The line of reasoning here is reminiscent of the much-discussed causal exclusion/closure argument against dualist and non-reductive physicalist views about the mind (e.g. Kim 1998). The argument goes, roughly, as follows. The physical domain is causally closed in that every physical effect (that has a cause at t) has a sufficient physical cause (at t). If some property P is not identical with some physical property, then, in order to cause some physical effect E, P (and some physical property) must causally overdetermine E. But such causal overdetermination is unacceptable, as it would imply systematic causal overdetermination. Therefore, the argument concludes, if mental properties are not physical properties, they must be epiphenomenal---they are causally excluded by physical properties. This argument is usually presented as a *reductio* against dualist and non-reductive physicalist views, as it allegedly corners such views into accepting an epiphenomenalist conclusion, which they typically want to avoid.

It is not clear to me that this argument is successful in forcing *nonphysicalist* dualists into an epiphenomenalist conclusion. Notwithstanding the controversial assumption that there is no systematic causal overdetermination, the denial of the causal closure of the physical is an open alternative for nonphysicalist dualists. This doesn't mean that nonphysicalists must deny the causal closure of the physical; but they *can*. In fact, it is not uncommon to find a nonphysicalist philosopher explicitly doing this. (See O'Connor 1994; Lowe 2013: 168-9; and to some extent, Chalmers 1996; 2003.) So, when looking for the link from nonphysicalism to QE, we shouldn't appeal to a principle that is questioned by nonphysicalists.

Is there another way to link the knowledge argument (or the associated thought experiment) to QE? Campbell (2003: 262), in arguing against Jackson's epiphenomenalist dualism, reconstructs what he takes to be Jackson's implicit argument. According to how Jackson expects us to understand this thought experiment, we must grant to Mary a *complete* understanding of all the physical causes and effects of colour stimulation on normal perceivers. This is meant to imply that she doesn't notice any mysterious causal gaps at the *physical* level. Given this *physical completeness* assumption, we are then asked to infer that there is no room for *additional* nonphysical causes.[2] But again, this is just another way of saying that unless qualia are physical properties the only way they can have physical effects would be by means of systematic causal overdetermination. However, if the choice is between QE and systematic causal overdetermination, it is not clear to me that the nonphysicalist has to choose QE (over causal overdetermination).

So, the knowledge argument and the associated thought experiment don't seem to give us QE, even if we were to accept them. What is worse is that there is at least an apparent tension

---

[1] It is worth emphasising that, unlike Jackson, Lewis doesn't think that that qualia are nonphysical or epiphenomenal, but he agrees with Jackson that *if* they are nonphysical, then they are epiphenomenal.
[2] Thanks to an anonymous referee here for helping me with this interpretation.

between the associated thought experiment and QE, as others have also observed (Watkins 1989; Campbell 2003). The tension is that the conclusion of the knowledge argument (that qualia are nonphysical properties) is supported by the supposition that Mary acquires knowledge of qualia (phenomenal knowledge) when she is released, and according to these authors, there is something mysterious (if not incoherent) about acquiring the knowledge of something which is epiphenomenal. Now, this criticism seems to assume a causal theory of knowledge, and it is questionable if this assumption is warranted.[3] In what follows, I shall try to steer clear of issues about theories of knowledge. But what I can't steer clear of is metaphysics, as under what conditions qualia, *qua* properties, should turn out epiphenomenal is ultimately a question about the metaphysics of properties.

III. Epiphenomenal Properties

Having seen that being nonphysical doesn't make qualia epiphenomenal, we are led to wonder what can make them so. We can't answer this before settling, more generally, what makes a property epiphenomenal.

Although 'epiphenomenalism' in philosophy of mind is often used to refer to the view that mental properties don't cause physical events (which may suggest that they can have nonphysical---perhaps mental---effects), an epiphenomenal property, more generally, can be understood as a property without any causal power. There is reason to formulate QE with this understanding of an epiphenomenal property, rather than one which is restricted to only *physical* effects. Suppose we understand QE as the view that qualia don't have causal powers to contribute to physical effects, but have causal powers to contribute to *nonphysical* mental effects. Call this RQE ('R' for 'Restricted). On the face of it, RQE is unmotivated. Think about putative possible nonphysical mental effects of qualia (on the assumption that RQE and nonphysicalism about the mind are true): a belief about one's own experience, or a desire to satisfy one's need for something. Assuming such mental states may have physical effects (because neither RQE nor, more generally, QE is committed to wholescale mental epiphenomenalism), from transitivity of causation, qualia will inevitably contribute to physical effects, contradicting the core thesis of RQE.

So, to begin with, I shall take the question of what makes a property epiphenomenal to be the question of what makes a property fail to have *any* causal power (rather than merely any *physical* causal power). The flipside of that question is that of what it is for a property to have some causal power. Only after settling under what conditions a property may be said to have *some* causal power can we settle the question of under what conditions it may *not*. In answering this question, the view I shall endorse is that properties, as abstract entities, don't *really* have causal powers, but their bearers (i.e. concrete particulars) *really* do have causal powers. According to this view, if there is a specific nomological link between having a property P and having some causal power C, then P can be said to have C (derivatively of its bearers). The nomological link in question is that of *nomological necessitation*. That is, if the possession of a property P nomologically guarantees the possession of a causal power C, then (and only then) C should figure in the causal profile of P. This implies, I think correctly, that properties have the same causal profiles in worlds that are governed by the same laws of nature, and that if laws of nature guarantee that the bearers of a property must have some causal power, that causal power should figure in the causal profile of that property (in worlds that are nomologically

---

[3] See Nagasawa (2010) for related discussion.

alike).[4] If laws of nature hold necessarily (rather than contingently), arguably, this claim of nomological necessitation will be equivalent to a claim of metaphysical necessitation.[5]

Appealing to these observations, we can thus characterise the notion of an epiphenomenal property, as a first approximation, as follows:

> (EP) A property P is epiphenomenal if and only if there is no causal power C such that it is nomologically necessary that all bearers of P have C.

That is, if it is nomologically possible for some bearers of P to share no causal power whatsoever, then, and only then, P is epiphenomenal. I believe that this is a plausible way to think about the notion of an epiphenomenal property. As realists about properties, we want properties to ground similarities of objects. Properties with causal powers should therefore ground the similarities of the causal powers of their bearers. And conversely, properties without causal powers should be those that do not ground any such causal similarities. I admit that there is quite a bit of a metaphysical baggage here, and it is predictable that, under different assumptions about properties, laws, and causation, epiphenomenal properties will be understood in different ways. It is not surprising that different metaphysical views about these concepts lead to different assessments of QE. I don't think there is any shame in this though, as even those who may disagree with these metaphysical views can find something interesting in what those views lead to in the assessment of QE.

In presenting EP above, I said that it was a first approximation. This is because it needs to be modified in order to avoid trivial results whereby no property is epiphenomenal. Assuming that there may be causal powers that are, as a matter of nomological necessity, shared by all concrete particulars, EP leaves all concrete properties (i.e. properties of concrete particulars) non-epiphenomenal. (I don't know what those causal powers would be, but it is epistemically possible that there may be some.) Now, that there are no epiphenomenal properties is a defensible claim, and in fact, it is what the Eleatic Principle would have us believe. But in discussing whether qualia are epiphenomenal or not, we shouldn't assume the Eleatic Principle. (How anticlimactic would it be to discover that qualia are not epiphenomenal because 'to be is to have causal powers'?) In resolving this, we can make a distinction between *first-class* and *second-class* causal powers. Second-class causal powers are powers that are shared by all nomologically possible concrete particulars. (They are second-class, not because they are less real or less fundamental, but because there isn't any privilege accrued by having them.) Any causal power that is not a second-class causal power is a first-class causal power. In line with this, we can restrict the account of epiphenomenal properties to first-class causal powers. So, epiphenomenal properties are those properties whose instantiations don't nomologically necessitate the possession of any given first-class causal power.

Perfectly unproblematic examples of epiphenomenal properties in this sense are tautological properties, such as being-red-or-being-non-red. Since everything has this property, it is

---

[4] This also allows for the possibility that different properties may share some causal powers. In particular when there are nomological necessitation relations between properties, on this view, the necessitated property 'inherits', so to speak, its causal powers from its bearers. As argued in Baysan (2016), I believe it is a virtue of this view that it makes certain cases of causal overdetermination intelligible and unproblematic.

[5] There is more to be said on these matters, but a proper treatment of these will take us too far afield. Elsewhere, I argue for this view by eliminating several alternatives (Baysan 2018; forthcoming). Apparent objections to this view (e.g. from co-extensional properties) are dealt with in Baysan (2018: 425-6).

possible to have instances of this property that share no causal similarities whatsoever.[6] Perfectly unproblematic examples of non-epiphenomenal properties in this sense would be first-class causal powers themselves. Such properties can be expressed by predicates of the form 'has the power to φ'.

The dichotomy between perfectly unproblematic cases of epiphenomenal properties and perfectly unproblematic cases of non-epiphenomenal properties seems to mirror the dichotomy between perfectly non-natural properties and perfectly natural properties. Now, there is something to be said that the latter is a false dichotomy, as a property can be natural without being perfectly natural (Lewis 1983). Likewise, a property can be non-natural without being perfectly non-natural. We might think that perfectly natural properties are those that ground perfect resemblances among instances, whereas natural, but not perfectly natural, ones are good indicators of resemblances among their instances, but do not ground perfect resemblances.

If the latter dichotomy is a false one, might the former be a false one too? That is, does it make sense to think of properties as more (or less) epiphenomenal than others? In what follows, I will consider the possibility that it does.[7] So, I propose that what I said above should be an account of *perfectly epiphenomenal* properties. Just as we might think that a property can be natural without being perfectly natural, we might want to say that a property can be epiphenomenal without being perfectly epiphenomenal. If the only causal powers that the possession of a property nomologically guarantees are relatively superficial and non-natural causal powers, it makes sense to think that the property in question is, epiphenomenal, but not perfectly so.

This is an explicit departure from what appears (to me) to be a widely shared intuition that being epiphenomenal is an 'all or nothing' matter. Let me anticipate the following objection. 'A property is either epiphenomenal or not-epiphenomenal. It is like being pregnant; either you are pregnant, or you are not; there is no middle ground!' I have two things to say in response to this objection. First, I am not denying the law of excluded middle; it is true that for any property P, either P is epiphenomenal, or P is not epiphenomenal. What I am suggesting is that among those properties which are epiphenomenal, some may be more epiphenomenal than others. Second, I would like to think that 'pregnant' and 'epiphenomenal' are interestingly different concepts. We have a good grasp of 'pregnant' (though we could do better!), and it would be odd to speculate from the armchair as to how we should understand this concept. But things are different for 'epiphenomenal'. This is a philosophical term of art, it has different meanings in the writings of different philosophers, and it surely isn't odd to speculate from the armchair as to how to understand it. The claim that properties could be more (or less) epiphenomenal than others shouldn't be dismissed offhand. In fact, although the phrases 'more epiphenomenal' and 'less epiphenomenal' may not have been used, the very idea of proposing to understand qualia as epiphenomenal with respect to physical effects but not with respect to nonphysical effects (as per RQE, discussed above) seems to be friendly to the idea that epiphenomenalness can come in degrees.[8] In the beginning of this section, I dismissed that idea when it comes to formulating QE. But nevertheless, the very existence of that idea goes to

---

[6] One might question the viability of this example by suggesting that such properties are not genuine properties. I want to sidestep this question because even if one were to think, understandably, that tautological properties are not genuine properties, that they are epiphenomenal *can* be serviceable for that conclusion.

[7] In previous work (e.g. Baysan 2018), I had not considered this possibility. It now appears to me that this possibility has important implications regarding QE (as I shall argue in Section VI).

[8] Thanks to an anonymous referee for pointing this out.

show that the suggestion that epiphenomenalness can come in degrees is at least intelligible. For what it is worth, in what follows, I will put this suggestion in use.

In mentioning relatively superficial and non-natural causal powers, I am thinking about cases where we would have to design idiosyncratic machines or come up with thought experiments to find causal powers that can be attributed to properties. Think about the property of being-a-red-bus-or-a-seven-inch-gecko. Are there any first-class causal powers that are shared by all nomologically possible bearers of this property? Well, if you design a machine that says 'forty two' whenever it detects a red bus or a seven-inch gecko, then all bearers of being-a-red-bus-or-a-seven-inch-gecko will have the causal power to make this machine say 'forty two'. The causal power in question is definitely not a second-class causal power. (It would be a very strange world if all concrete particulars had this causal power!) But there is a sense in which this causal power is quite fishy. In fact, let's call such relatively superficial and non-natural causal powers 'fishy'. Fishiness is then a *limit* case of being a causal power of a property. Unfortunately, any attempt to give a more precise formulation of what makes a causal power fishy (i.e. relatively superficial and non-natural) seems to fail. I invite the reader to reflect on what kinds of causal powers she is happy to take as natural and which ones she is not, and take such intuitions on board. In sum, I am tempted to think that being-a-red-bus-or-a-seven-inch-gecko is an epiphenomenal property, but not a perfectly epiphenomenal property because it nomologically guarantees the possession of only fishy first-class causal powers.

IV. Epiphenomenal Qualia

When introducing the question of QE above, I suggested that we can assume that SQE is false: experiences are not epiphenomenal. We can debate whether this is a defensible claim, but it is important to note that the question of QE is a separate one. Experiences are instantiations of mental properties, and such properties are properties of persons, organisms, and suitably complex systems which can have relatively rich mental lives. (Henceforth, I will stick to 'persons' for ease of expression, but I don't mean this to suggest that only persons can have mental properties.) So, the bearers of such mental properties are persons. In holding that mental properties of persons are not epiphenomenal, one should not thereby be committed to the claim that qualia are not epiphenomenal, as the bearers of qualia are not persons, but mental states of such persons. Just to emphasise: the bearers of qualia are experiences, not persons. If mental properties of persons indeed have causal powers, that doesn't mean that *the properties of their instances* must have causal powers too.

Recall that in order to decide whether a property is epiphenomenal, we must consider the causal powers of the bearers of that property. The bearers of qualia are experiences. So, to examine whether a given quale is epiphenomenal, we must look at the causal powers of the experiences that instantiate that very same quale. On the account proposed here, *an individual quale Q is perfectly epiphenomenal just in case it is nomologically possible that there are experiences that instantiate Q which share no first-class causal powers in common*. That is, if it is nomologically possible that there are experiences that instantiate the same quale but share no first-class causal powers, then (and only then) that quale is perfectly epiphenomenal. Let's now qualify QE to be the view that qualia are perfectly epiphenomenal.

Do we have good reasons to think that qualia are perfectly epiphenomenal in light of this criterion? What sorts of cases (or the nomological possibility thereof) might support QE? In exploring this, I will consider cases of absent qualia, inverted qualia, and (what I shall call) mad qualia. As I will argue---contrary to what may appear to be a natural thought---cases of absent and inverted qualia are irrelevant to the question of QE.

V. Absent Qualia

Absent qualia cases are traditionally discussed in relation to functionalism about mental properties (see Block & Fodor 1972; Block 1980; Shoemaker 1975; 1982). Suppose that the causal profile of being in pain is such that its instantiation is caused by inputs $I_1, I_2 \dots I_n$, causes states $S_1, S_2 \dots S_n$ and behavioural outputs $O_1, O_2, \dots O_n$. According to functionalism, necessarily, a person is in pain just in case she is in a state that meets this causal profile. Suppose that there is a person, Zed, who is in a state that is caused by inputs $I_1, I_2 \dots I_n$, causes states $S_1, S_2 \dots S_n$ and behavioural outputs $O_1, O_2, \dots O_n$. But for Zed, there is nothing that it is like to be in this state. Now consider Ned, who is in the same causal/functional state as Zed, but for Ned, there is something that it is like to be in this state. Whereas Zed's mental state instantiates no qualia, Ned's mental state instantiates a quale, say, painfulness. (Ned's situation should be familiar to anyone who has felt pain.) In the terminology employed here, whereas Ned's mental state is an experience, Zed's is not.[9] Ned and Zed's case is a typical case of absent qualia: functionally isomorphic states of two individuals, one state instantiates a quale, the other doesn't. With the assumption that pain couldn't be pain without being painful, it is arguable that if Zed's case is metaphysically possible, then functionalism about pain is false.

It is natural to think that if there could be cases of absent qualia, then qualia must be (perfectly) epiphenomenal, as it doesn't look like Ned's mental state's instantiation of painfulness is adding any causal powers to that mental state. We know from Zed's case what causal profile that mental state would have had if it hadn't instantiated a quale: the very same one that Ned's mental state has. As natural as this thought might be, it is unfortunately a misleading one. Based on the account of epiphenomenal properties employed here, painfulness is perfectly epiphenomenal just in case it is nomologically possible that there are painful experiences that share no first-class causal powers in common. So, to test whether a quale is epiphenomenal, we need two (or more) experiences with the same quale. But if qualia are absent from a mental state, that mental state is not an experience. Therefore, absent qualia cases, even if they are nomologically possible, don't give us cases to examine QE. This is an interesting result that is worth paying attention to---and perhaps worth considering an objection against.

Here is one objection. If, on my view, the possibility of absent qualia doesn't show that qualia are (perfectly) epiphenomenal, then my view must be mistaken. Let's see what is guiding this objection. Setting aside qualia momentarily, consider the following case. Suppose we have two objects, a and b, with precisely the same causal powers C_1, C_2, C_3, and C_4. Moreover, a and b have also the same properties, P_1, P_2, P_3, P_4, *with only one difference*: a also has the further property Q, whereas b lacks Q. On the face of it, the possibility of this case seems to suggest that Q is an epiphenomenal property. Having Q isn't making any difference with respect to the causal powers of a (just as having painfulness is not making any difference to the causal powers of Ned's mental state, if Zed's case is possible). But I think we should resist this conclusion. The case is very abstract, so it is hard to imagine what the relevant causal powers of Q might be, but it can be shown schematically that it is possible that although Q doesn't make a difference to the causal powers of a, it may not be epiphenomenal as it may happen to confer C_1 (or any other causal power that is conferred by P_1, P_2, P_3, or P_4). There is no rule that the very same causal power can't be conferred by different properties. I have the causal power to

---

[9] As an anonymous referee rightly points out, it is possible for Zed's mental state to be an experience even if it fails to instantiate painfulness, as it should possible that one experience can have numerous qualia. For this thought experiment to depict a case of absent qualia (so that we can assess the thought experiment), I will assume that the counterpart experience of Ned instantiates only painfulness and no other qualia.

reach my toes. I can manifest this power by using either my left hand or my right hand. That I can exercise this power by using my right hand doesn't mean that my having a left hand is not conferring me the causal power to reach my toes.

VI. Inverted Qualia

Inverted qualia cases, like absent qualia cases, are also widely discussed in relation to functionalism. In an inverted qualia case we have two persons, let's say Ned and Ted, whose mental states don't differ in their causal profiles but differ in their qualia. Let's say that both Ned and Ted are having visual experiences upon looking at first a green screen, and then a red screen. Let's suppose that they are disposed to form the same beliefs and elicit the same behavioural responses upon these two experiences. But, let's imagine that what it is like for Ned to see something red is different from what it is like for Ted to see something red, and in fact identical to what it is like for Ted to see something green. And likewise, what it is like for Ned to see something green is different from what it is like for Ted to see something green, and in fact identical to what it is like for Ted to see something red. In other words, Ned and Ted's red and green visual experiences are qualitatively inverted.

It is less obvious (if obvious at all) that if inverted qualia are metaphysically possible, functionalism must be false. Shoemaker (1975; 1982), while defending functionalism, admits that inverted qualia cases are metaphysically possible, but goes on to argue that similarity and difference relations between qualia can be causally/functionally characterised. In order for Ned and Ted's qualia inversion to be functionally indiscernible, Ned and Ted must be making the same similarity and difference judgements regarding colours. That is, if asked whether orange is more similar to red than green, both should give the same response (which goes to suggest that, for functionally indiscernible colour qualia inversion, we need the inversion of the entire colour spectrum, not only a pair of colour qualia). If this is true, although individual qualia cannot be given causal/functional characterisations, their similarities and differences can be. Or so argues Shoemaker. Again, my concern is not whether functionalism survives the possibility of qualia inversion. I am interested in whether there is an argument from the possibility of inverted qualia to epiphenomenal qualia.

In an inverted qualia case, we seem to have two discrepancies (relative to normal cases). First, we have different qualia accompanying experiences with identical causal profiles. Different qualia are instantiated by Ned's visual experiences of seeing something red and Ted's visual experiences of seeing something red, but Ned's and Ted's visual experiences of seeing something red have identical causal profiles. Second, we have the very same quale accompanying experiences with different causal profiles. The quale that accompanies Ned's visual experiences of seeing something red is identical in type with the quale that accompanies Ted's visual experiences of seeing something green, but the visual experience of seeing something red and the visual experience of seeing something green have different causal profiles. One is caused by red objects and causes beliefs and verbal reports about red objects; the other is caused by green objects and causes beliefs and verbal reports about green objects. The first of these discrepancies, like in the case of absent qualia, is irrelevant to the question of QE. For qualia to be epiphenomenal, we need the possibility of the same quale being instantiated by experiences that have different causal powers. But the second discrepancy gives us such a possibility.

Now, the second discrepancy gives us a case of qualitatively identical experiences with different causal powers; but it doesn't give us a case of experiences with *entirely* different causal powers. Arguably, despite differing in some of their causal powers, the visual experience

of seeing something red and the visual experience of seeing something green will share some causal powers too. For example, both a red and a green chip will stand out in a bunch of white chips, suggesting that the experiences caused by them will share some causal powers, e.g. the causal power to discriminate the odd chip. This goes to show that even if inverted qualia were nomologically possible, such possibility would not show that qualia are *perfectly* epiphenomenal; their bearers (experiences) share some first-class causal powers.

But I have argued that a property may be epiphenomenal without being perfectly epiphenomenal. If the only causal powers that are shared by Ned's and Ted's experiences that bear the same quale are fishy (i.e. they require idiosyncratic machines or thought experiments to be discovered), then the quale in question can be epiphenomenal without being perfectly epiphenomenal. Does the imagined scenario of inverted qualia show that the relevant experiences share only fishy causal powers? I don't think so. The causal power of an experience that enables its subject to discern objects that stand out among other objects doesn't seem to be a fishy causal power. We don't need to imagine a farfetched thought experiment to unearth such a causal power. It is exactly the type of causal power that we would naturally attribute to visual experiences (if we wanted to do so). So, the case we have considered doesn't give us epiphenomenal qualia, perfectly or not perfectly.

Perhaps we can imagine cases where experiences that instantiate the same quale are as different from each other as two states could be (with respect to their causal powers). But for that, we don't really need inverted qualia. We need mad qualia.

VII. Mad Qualia

Recall Lewis's story of the madman. The madman's pain feels exactly like our pain. (I am assuming that we are 'typical' subjects of experience.) Our pain and the madman's pain are qualitatively identical; they instantiate the very same quale. But the madman's pain has an entirely different causal profile than our pain.[10] (See the opening passage above.[11]) The madman's case is a case of mad qualia. More generally, mad qualia cases are cases where the very same quale is instantiated by experiences that have entirely different causal profiles. It follows from the characterisation of what makes a quale perfectly epiphenomenal that if mad qualia are nomologically possible, then (and only then) qualia---or at least those qualia which can 'go mad' (i.e. those qualia for which mad qualia cases are possible)---are perfectly epiphenomenal.

It looks like we finally have a test for perfectly epiphenomenal qualia: find some qualia that can go mad, and then establish QE for at least those qualia; make an argument that any quale can go mad, establish QE once and for all. But that seems to shift the question to the following: *What does it take for mad qualia to be nomologically possible?* Focusing on the case of mad pain, let us now consider this question. For the madman's case, we need two central ingredients. First, the madman's experience has to be a pain experience. Second, that experience has to have the strange causal profile that the story tells us. Let's focus on the first of these ingredients.

---

[10] It is important to note that causal powers can be dispositional. In a stronger reading, the madman's pain is not even disposed to bring about $S_1, S_2 \ldots S_n$ and $O_1, O_2, \ldots O_n$ in response to $I_1, I_2 \ldots I_n$.

[11] To be precise, the opening passage doesn't tell us that normal pain and mad pain have *entirely* different causal profiles. As far as that passage goes, normal pain and mad pain may have grossly-but-not-entirely different causal profiles. However, I will be assuming henceforth that normal pain and mad pain have entirely different causal profiles.

What does it take for the madman to be in pain? Or indeed, what does it take for any person to be in pain? According to Lewis (1972; 1989), for a person to be in pain, she needs to be in the state that *realizes* being in pain in her species (or in a relevant population). The relevant sense of realization here is that of occupying a causal role. The typical causal role of pain, remember, is that it is caused by inputs $I_1, I_2 \ldots I_n$, it causes states $S_1, S_2 \ldots S_n$ and behavioural outputs $O_1, O_2, \ldots O_n$. Let's give in to the philosophical fiction that being in C-fibre stimulation (C-fs for short) is the state that is caused by inputs $I_1, I_2 \ldots I_n$ and that causes states $S_1, S_2 \ldots S_n$ and behavioural outputs $O_1, O_2, \ldots O_n$ in humans. In other species, for example in some pain-capable Martian species, different physical states may play this causal role. Now, the madman is a human. So, in order for the madman to be in pain, he has to be in a state that occupies the relevant causal role: being caused by inputs $I_1, I_2 \ldots I_n$, causing states $S_1, S_2 \ldots S_n$ and behavioural outputs $O_1, O_2, \ldots O_n$ in humans. But we also know that the state that he is in is *not* caused by $I_1, I_2 \ldots I_n$; nor does it cause states $S_1, S_2 \ldots S_n$ and behavioural outputs $O_1, O_2, \ldots O_n$. How can this be? Lewis's answer is that the madman is indeed in that very state, which we agreed is C-fs. He is in pain because *he is in the state that realizes pain in his population*; but 'he is hooked up wrong' (Lewis 1980: 219).[12] That is, he is in C-fs, and C-fs indeed has the aforementioned causal profile; but there is something wrong in the madman's neural circuitry such that *in him*, C-fs is caused by moderate exercise on empty stomach and causes leg-crossings and finger-snappings.

There is something wrong with this story. As Shoemaker (1981: 97) points out, there are two senses in which one can be said to be in a state that realizes pain (or any mental property for that matter). First, one can be in a physical state which is a *core* realizer of being in pain. Call this core realizer state CR. Whether a human-person is in pain depends on whether she is in CR, and her being in CR is 'in some sense constitutive of [her] being in pain' (ibid.). But this will be so only if she is also in a *total* realizer state for being in pain---and this is the second sense in which one can be said to be in a state that realizes pain. Call this total realizer TR. This state TR will be a state that includes, alongside CR, the relevant background condition that is expressed by 'being physically constituted in such a way that CR plays the causal role of being caused by inputs $I_1, I_2 \ldots I_n$, causing states $S_1, S_2 \ldots S_n$ and behavioural outputs $O_1, O_2, \ldots O_n$'. Now, although being in CR will not be sufficient, metaphysically or nomologically, for being in pain (because it will not be sufficient for being in a state that occupies the relevant causal role), being in TR will be.

What is wrong in Lewis's case of the madman is that the madman is clearly not in a total realizer state for being in pain. He is in C-fs alright, and if the relevant background conditions had been present, his being in C-fs would have constituted his being in pain. But the madman is 'hooked up wrong', hence the relevant background conditions are not there. If our sole reason for thinking that the madman is in pain is that he is in the appropriate physical state for being so, since he is not in that physical state, it is odd to think that he is in pain.

In fact, if being in pain is a matter of being in some appropriate physical state that plays a causal role, the madman's case shouldn't be nomologically possible. Let me put this point in terms of a dilemma: either the madman has a total realizer of pain or he doesn't. If he does, then he is in pain, but the core realizer of his pain indeed plays the causal role that is characteristic of pain. If he doesn't, then the core realizer doesn't play this causal role (and perhaps it is caused by moderate exercise on empty stomach and causes leg-crossings and finger-snappings), but

---

[12] With this 'hooked up wrong' comment, differently from the classic absent and inverted qualia cases, we have an explanation of how the relevant experience is detached from its characteristic causal role. Thanks to an anonymous referee for pointing this out.

he isn't in pain. All this seems to suggest that if the madman's case is nomologically possible, then there is more to being in pain than being in some physical state that plays some causal role. And more generally, if mad qualia are possible, experiences seem to be 'over and above' physical states.

Here is a caveat, which I must concede. Surely, just because the particular physicalist assumptions considered in the previous four paragraphs can't account for the nomological possibility of mad qualia, it shouldn't thereby follow that no physicalist theory can do this. But what we have seen seems to suggest that the nomological possibility of mad qualia is in tension with the dominant variety of non-reductive physicalism, namely *realization physicalism* (e.g. Melnyk 2003; Shoemaker 2007; Wilson 2011). Also, recall that, because of the possibility of Martian pain, we have already ruled out reductive physicalism (a.k.a., the type-identity theory). Realization physicalism makes it a requirement that all mental properties are physically realized. Although the details of how 'realization' should be understood varies from one theorist to another, a shared commitment is that the causal profile of a realized property is an essential ingredient in what it takes to realize it.

To expand on this last claim, consider the debate between Gillett's (2002; 2003; 2010) so-called *dimensioned* view of realization and what he calls the *flat* views of realization. According to the flat views, a realized property and its realizer are instantiated by the same object, and according to Gillett, such views endorse the claim that the causal powers of a realized property are also the causal powers of its realizer.[13] Gillett's dimensioned view on the other hand rejects this picture. Instead, Gillett argues that realizing properties (and relations) are instantiated by the constituent parts of the object which instantiates the realized property. If this is true, Gillett argues, the causal powers of a realized property will not typically be had by the causal powers of the realizer properties, as the latter will be properties of micro objects whereas the former will be properties of macro objects. But Gillett still takes the causal powers of realized and realizer properties to be essential to his account of realization: higher-level properties of macro objects are ('dimensionally') realized by lower-level properties of micro-objects because higher-level properties confer their causal powers on macro objects in virtue of the conferral of the causal powers by lower-level properties on micro-objects. Another realization theorist who rejects the flat view (without employing Gillett's terminology) is Pereboom (2002; 2011). But again, the causal powers of realized properties are an essential component of Pereboom's account of realization: higher-level properties are realized by lower-level properties because the causal powers of the former are 'constituted' by the causal powers of the latter. What this constitution comes down to is not relevant to our purposes here. What is important is the observation that virtually all prominent accounts of realization attempt to explain realization partially in terms of the causal powers of realized properties.[14]

Now, if realization physicalism is true and hence all experiences must be physically realized, and moreover that the realization relation must take into account the causal profile of a realized property, then it would be very mysterious, from a realization physicalist point of view anyway, for an experience type to have no robust causal profile. If experiences are not type-identical with physical states and also are over and above physical states that occupy causal roles, what could explain the fact that two experiences, E_1 and E_2, are instances of the same experience type (e.g. being in pain)? The only option would be to appeal to some brute phenomenal aspect of those experiences: E_1 and E_2 are instances of the same experience type because they *feel* the same, and nothing other than this constitutes the fact that they are instances of the same

---

[13] Gillett attributes the flat theory to Shapiro (2000), Clapp (2001), Wilson (2009).

[14] See Baysan (2015) for a more detailed discussion of the aforementioned realization theories.

experience type. So, seemingly, in order to account for the possibility of mad qualia, we need to appeal to brute phenomenal facts, and such facts appear to fly in the face of physicalism.

VIII. Conclusions

I have characterised mad qualia cases as cases in which the very same quale is instantiated by experiences that have entirely different causal profiles. I then argued that if being in pain consists in having some physical property that plays some causal or functional role, then mad pain is not nomologically possible. And more generally, if having an experience (i.e. having a mental state that instantiates a quale) consists in having some physical property that plays some causal or functional role, then mad qualia are not nomologically possible. If physicalism is true, then mad qualia shouldn't be nomologically possible. To contrapose: if mad qualia are nomologically possible, then physicalism should be false. This is what we can tentatively conclude from the discussion in §VII.

I have also argued that a quale is perfectly epiphenomenal if and only if it is nomologically possible that there are experiences that instantiate this very same quale but they are entirely different in their causal profiles. This was the conclusion of §IV. From the characterisation of mad qualia, it follows that qualia are perfectly epiphenomenal just in case mad qualia are nomologically possible. More precisely, a quale is perfectly epiphenomenal just in case that quale can 'go mad' (i.e. a mad quale case is possible for it). From this and the conclusion of §VII, it follows that if qualia are perfectly epiphenomenal, then physicalism should be false.

This is not a trivial result---though I can imagine that it might strike some as one. This would be a trivial result if it were established that all and only nonphysical properties are epiphenomenal. But this is not established. In §II, I sought an argument from qualia nonphysicalism to QE, and concluded that a property's being nonphysical doesn't thereby constitute grounds for its being epiphenomenal. That there are nonphysical yet non-epiphenomenal properties is an open alternative---and it is one that nonphysicalist philosophers can and do appeal to. Moreover, and more interestingly, that there are physical yet epiphenomenal properties is also an open, but underexplored, alternative. First, it is not part of the definition of 'physical' that something physical must have causal powers. Second, from the account of epiphenomenal properties that I have appealed to in this paper, it seems possible that there can be physical properties that are perfectly epiphenomenal. A property is perfectly epiphenomenal if and only if it is nomologically possible that its bearers share no first-class causal powers. Now, suppose that there is a disjunctive property that consists of the disjunction of all nomologically possible physical properties. This property will be physical, as it has only physical properties as constituents; but it has a very good claim to be perfectly epiphenomenal because there is no first-class causal power that is nomologically guaranteed by the possession of this property. Now, I am not sure if we should admit such a property in our ontology. But if we shouldn't, I would like to think that it has something to do with the fact that it is a perfectly epiphenomenal property. Either way, the example is serviceable for the claim that the concept of a 'physical property' doesn't entail that a physical property must be a non-epiphenomenal one.

I don't know if mad qualia are nomologically possible or not. Likewise, I don't know if physicalism about qualia is true or not. But what we have seen suggests that qualia are epiphenomenal if and only if mad qualia are nomologically possible, and the possibility of mad qualia is in tension with physicalism. Although nonphysicalism and epiphenomenalism about qualia have long been strongly associated, I think it wasn't clear in what ways these two positions were really connected, and what could ground this association. I believe I have shown

how these views are linked and hopefully offered a framework that can explain their association.[15]

References

Baysan, U. (2015) 'Realization Relations in Metaphysics', *Minds and Machines* 25/3: 247–260.

Baysan, U. (2016) 'An Argument for Power Inheritance', *The Philosophical Quarterly* 66: 383–390.

Baysan, U. (2018) 'Epiphenomenal Properties', *Australasian Journal of Philosophy* 96/3: 419-43.

Baysan, U. (forthcoming) 'Causal Emergence and Epiphenomenal Emergence', *Erkenntnis*: 1-14.

Block, N. J., & Fodor, J. A. (1972) 'What Psychological States are not', *The Philosophical Review*, 81/2: 159-181.

Block, N. (1980) 'Troubles with Functionalism', in N. Block, (ed.) *Readings in Philosophy of Psychology Volume 1*, 268-305. Cambridge, MA: Harvard University Press.

Campbell, N. (2003) 'An inconsistency in the Knowledge Argument', *Erkenntnis*, 58/2: 261-266.

Chalmers, D. J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, D. J. (2003) 'Consciousness and its Place in Nature', in S. Stich & T. Warfield (eds.) *Blackwell Guide to the Philosophy of Mind,* 102-142. John Wiley & Sons.

Clapp, L. (2001) 'Disjunctive Properties', *Journal of Philosophy*, 98: 111–136.

Gillett, C. (2002) 'The Dimensions of Realization: A Critique of the Standard View', *Analysis*, 62: 316–323.

Gillett, C. (2003) 'The Metaphysics of Realization, Multiple Realizability and the Special Sciences', *Journal of Philosophy*, 100: 591–603.

Gillett, C. (2010) 'Moving beyond the Subset Model of Realization: The problem of Qualitative Distinctness in the Metaphysics of Science', *Synthese*, 177: 165–192.

Jackson, F. (1982) 'Epiphenomenal Qualia', *The Philosophical Quarterly*, *32*: 127-136.

Kim, J. (1998) *Mind in a Physical World: An Essay on the Mind-body Problem and Mental Causation*. Cambridge, MA: MIT press.

Lewis, D. (1972) 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, 50/3: 249-258.

Lewis, D. (1980) 'Mad Pain and Martian Pain', in N. Block, (ed.) *Readings in Philosophy of Psychology Volume 1*, 216-222. Cambridge, MA: Harvard University Press.

Lewis, D. (1983) 'New Work for a Theory of Universals', *Australasian Journal of Philosophy*, 61/4: 343-377.

Lewis, D. (1988) 'What Experience Teaches', *Proceedings of the Russellian Society*, 13: 29–57. Reprinted in P. Ludlow et. al. (eds.)*, There's Something about Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, 77-103. Cambridge MA: MIT Press, 2004.

Lowe, E. J. (2013) 'Substance Causation, Powers, and Human Agency', in S.C. Gibb et. al. (eds.) *Mental Causation and Ontology*, 153-172. Oxford: Oxford University Press.

McLaughlin, B. (1989) 'Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical', *Philosophical Perspectives*, 3: 109-135.

Melnyk, A. (2006) 'Realization and the Formulation of Physicalism', *Philosophical Studies,* 131: 127–155.

Nagasawa, Y. (2010) 'The Knowledge Argument and Epiphenomenalism', *Erkenntnis*, 72/1: 37-56.

O'Connor, T. (1994) 'Emergent Properties', *American Philosophical Quarterly*, 31/2: 91-104.

Pereboom, D. (2002) 'Robust Nonreductive Materialism', *Journal of Philosophy*, 99: 499–531.

Pereboom, D. (2011) *Consciousness and the Prospects of Physicalism*. New York, NY: Oxford University Press.

Shapiro, L. (2000) 'Multiple Realizations', *Journal of Philosophy,* 97/12: 635-654.

Shoemaker, S. (1975) 'Functionalism and Qualia', *Philosophical Studies*, 27/5: 291-315.

Shoemaker, S. (1981) 'Some Varieties of Functionalism', *Philosophical Topics*, 12/1: 93-119.

Shoemaker, S. (1982) 'The Inverted Spectrum', *The Journal of Philosophy*, 79/7: 357-381.

Shoemaker, S. (2007) *Physical Realization*. Oxford: Oxford University Press.

Watkins, M. (1989). 'The Knowledge Argument against "The Knowledge Argument"', *Analysis,* 49/3: 158-160.

Wilson, J. (2009) 'Determination, Realization, and Mental Causation', *Philosophical Studies*, 145: 149–169.

Wilson, J. (2011) 'Non-Reductive Realization and the Powers-Based Subset Strategy', *The Monist*, 94: 121–54.

Oxford University, UK