

# AN ARGUMENT AGAINST EPIPHENOMENALISM

JASON MEGILL

Carroll College

## ABSTRACT

---

I formulate an argument against epiphenomenalism; the argument shows that epiphenomenalism is extremely improbable. Moreover the argument suggests that qualia not only have causal powers, but have their causal powers necessarily. I address possible objections and then conclude by considering some implications the argument has for dualism.

**Keywords:** Epiphenomenalism; Dualism; Mind-Body Problem; Supervenience

## 1. Epiphenomenalism

Epiphenomenalism claims that although mental events are caused by physical events, mental events do not causally affect the physical in turn (e.g., see Heil 1998, 37-38 or Robinson 2003). This paper concerns the causal efficacy of qualia; it will not address issues concerning the causal efficacy of beliefs. So, for the remainder of this paper, ‘epiphenomenalism’ refers to the view that qualia specifically do not affect the physical. ‘Qualia’ is used in the usual sense; qualia are mental states “with a very distinctive subjective character...There is something it is like for me to undergo each state;” qualia are “introspectively accessible, phenomenal aspects of our mental lives” (Tye 2013, introduction). In this section, I argue that qualia are not epiphenomena.

It is clear that certain types of bodily states are associated with or paired with certain types of qualia. For example, there is a close connection between tissue damage and pain qualia: where there is tissue damage, there is often pain. Different views in philosophy of mind will offer different accounts of such regularities; e.g., the identity theory will claim that tissue damage causes brain states that the qualia are identical to in some sense, while dualism will claim that the bodily states cause brain states which in turn cause the qualia. But for now, simply note such regularities, whatever the true explanation of them might be. Likewise, there are close connections between certain types of

qualia and certain types of behavior. For example, there is a connection between pain and avoidance behavior: where there is pain, there is often avoidance behavior. It is this latter type of regularity, i.e., the connection between certain types of qualia and certain types of behavior, that is most relevant for our purposes. It is undeniable that such associations or pairings exist; and to be clear, when it is claimed that certain types of qualia are often “associated with” or “paired with” certain types of behavior, this simply means that the two generally appear together. Of course, these associations only hold most of the time, e.g., pain medication might keep pain qualia away in the face of tissue damage, but this sort of example will not affect the argument.

Given that such associations exist, there is one aspect of them in particular that will play a role in the argument. To be specific, these associations do not appear random. Rather, at least in some cases, given the subjective nature of a given type of qualia, the behavior that is associated with the qualia seems “appropriate.” That is, given the intrinsic natures of some types of qualia, the behavior that is associated with the qualia is precisely what one would expect. There is often a high degree of “fit” or “appropriateness” to these associations. Qualia and behavior often seem to be “aligned” in the correct manner. To offer an example, again, pain qualia are generally associated with avoidance behavior. Pain qualia hurt, they are subjectively unpleasant. Given this, avoidance behavior is exactly the sort of behavior one would expect to be associated with pain qualia. To offer a different example: consider someone who displays avoidance behavior when they have blue qualia. This strikes us as extremely bizarre precisely because avoidance behavior is not an appropriate behavior given blue qualia. There is nothing about the intrinsic nature of a blue quale that would suggest that it should be paired with (or correlated with) avoidance behavior. In short, the claim is the following: qualia have intrinsic natures. Some qualia, such as pain, are intrinsically unpleasant; they are unpleasant in and of themselves; they are unpleasant even in isolation from anything else (such as any beliefs we might have). Other qualia, such as pleasure, are intrinsically pleasant; they are pleasant in and of themselves; they are pleasant even in isolation from anything else. Moreover, given the intrinsic nature of these qualia, we would expect them to be paired with certain behaviors and not paired with other behaviors. We would (and do) expect pain to be correlated with avoidance behavior (and it is). We would (and do) expect pleasure to not be correlated with avoidance behavior (and it is not). We are surprised when these expectations are violated, as they are in, e.g., cases of masochism (which is discussed below).<sup>1</sup>

I now formulate the argument. Of course, qualia either have causal powers or they do not; epiphenomenalism is either true or false. And again, there is a certain fit between the subjective natures of some types of qualia and the types of behavior that are associated with the qualia; given the nature of the qualia, the behavior seems appropriate. Suppose that qualia do not have causal powers, i.e., epiphenomenalism is true. But this is likely false for the following reason: if epiphenomenalism is true, then the “fit” between qualia and behavior is quite remarkable. Indeed, it would be nothing short of miraculous. For if

---

<sup>1</sup> One might think that qualia do not have “intrinsic natures” or “subjective natures.” But note that dualists generally do think that qualia have intrinsic and/or subjective natures, and epiphenomenalism is a prominent form of dualism; so dualistic epiphenomenalism, at least, cannot and will not deny these claims to avoid the argument.

qualia are epiphenomena, they play no role in the production of behavior; after all, they do not causally influence the physical. But then any type of qualia could, in principle, be associated with any type of behavior. For example, happiness might as well be associated with avoidance behavior: it wouldn't make any difference. Therefore, given this, the fit between qualia and behavior is indeed quite surprising, so surprising, in fact, that it calls our assumption, i.e., epiphenomenalism, into question. So, epiphenomenalism is very improbable.<sup>2</sup> Indeed, there is a far more plausible view: qualia and behavior show a high degree of "fit" because qualia *cause* behavior. The reason that pain qualia are associated with avoidance behavior is that pain hurts, pain is unpleasant, and it causes us to recoil. This is the common sense view, but it is a view that is not taken seriously enough in contemporary philosophy of mind.<sup>3</sup> So, we seem to be faced with the following choice: we can endorse the common sense view, or we can endorse a view that is highly improbable; it seems we should opt for the former. In effect, we should make an "inference to the best explanation" (see Harman 1965).

To offer a slight reformulation of the argument: there are various types of qualia and there are various types of behavior. Assume that epiphenomenalism is true. If so, there is no causal connection between qualia and behavior; the subjective nature of a given qualia type does not shape behavior in any way. But then, in theory, any given type of qualia could be associated with – or paired with – any given type of behavior: qualia and behavior are independent of one another, at least in the causal sense. Further, there are very many possible pairings or associations when it comes to qualia and behavior. For example, avoidance behavior might have been associated with happiness, or blue, or pain, or any other type of qualia. Some of these pairings will exhibit a certain degree of appropriateness, e.g., those in which pain is associated with avoidance behavior, but a vast majority will not, e.g., those in which happiness, or blue, is associated with avoidance behavior. So, given epiphenomenalism, it is highly improbable that we would find ourselves in a world in which qualia and behavior show a high degree of "fit" (again, there are far more inappropriate associations than appropriate ones). But we are in such a world, so it is very likely that epiphenomenalism is false.

## 2. Further Implications

The claims in section one suggest that two additional claims are true. First, the argument shows that, at least in some instances, different types of qualia have different causal powers. Again, the argument claims that there is a certain degree of fit between qualia and behavior, and this fit suggests that qualia cause behavior (for otherwise this fit would be an incredible coincidence). Further, suppose that different types of qualia all have the same

---

<sup>2</sup> This argument shares a premise with one offered by W. James (1879; 1890); both arguments appeal to the appropriate alignment between qualia and other factors such as behavior to call epiphenomenalism into question. There are significant differences between this argument and James's argument as well, e.g., James made crucial appeal to natural selection in his argument (see Robinson (2003) for a brief discussion of James's argument). (If the reader is unfamiliar with James's argument, it was essentially the following: if qualia are epiphenomena, then we might as well enjoy harmful things and avoid things that aid our existence (but we do not). And if qualia are epiphenomena, natural selection cannot explain the appropriate alignments between beneficial things and what we enjoy, so we would have to take this alignment as a brute fact or a sort of pre-established harmony.)

<sup>3</sup> To explain, many physicalists, at least, do not think that the subjective nature of a quale has causal powers. Indeed, some physicalists deny there are qualia (see, e.g., Dennett 2005). Likewise, epiphenomenalism is of course a prominent form of dualism (e.g., Kim (2005) claims that the subjective natures of qualia are non-physical and epiphenomenal).

causal powers. So, e.g., happiness, or any other type of qualia, for that matter, could cause avoidance behavior. But if so, then we are once again faced with a highly improbable coincidence. If any type of qualia can cause avoidance behavior, then it is surprising that pain, as opposed to some other type of qualia, is associated with avoidance behavior. In other words, it is surprising that there is a high degree of fit between qualia and the behavior that the qualia causes, if indeed another type of qualia could have caused the behavior just as well. In effect, if different types of qualia cannot have different causal powers, then the appropriate alignment between qualia and behavior becomes incredible once again. Therefore, it must be that different types of qualia can have different causal powers.

Second, and this is somewhat surprising, the key intuition on which the argument rests suggests that qualia have their causal powers necessarily, or in all metaphysically possible worlds. The reason this is surprising is that philosophers often think the laws of nature are metaphysically contingent. Indeed, if entities have their causal powers in all metaphysically possible worlds, it seems the commonplace distinction between nomological and metaphysical necessity collapses. Again, the key intuition on which the argument rests is that given the subjective natures of certain types of qualia, the behaviors that are associated with the qualia are appropriate; and this intuition suggests qualia cause behavior. In other words, the subjective nature of a quale determines its causal powers: the quale has the causal powers that it has precisely because of its subjective nature. The essence of a quale, so to speak, determines its causal powers. But then, if the particular causal powers of qualia are fixed by the very nature of the qualia, anytime the qualia are present, the particular causal powers will be present as well. In effect, qualia will have their causal powers necessarily, or in all metaphysically possible worlds. To elaborate, the “fit intuition” suggests that there is a connection between the subjective nature or essence of a quale and its particular causal powers; the former determines the latter. But a quale will have the same subjective nature or essence in any metaphysically possible world in which it exists (this seems trivially true: if a quale has a different subjective nature from pain, say, if it feels pleasant, then it is necessarily not pain). But then, since the essence of a quale determines its causal powers, a quale will have the same causal powers in all metaphysically possible worlds.<sup>4</sup>

### 3. Possible Objections

I now address possible objections.

First, one might object that the argument proves little; after all, even if epiphenomenalism is highly improbable, it still might be true nevertheless. However, this objection should not be made lightly; it seems to underestimate just how highly improbable epiphenomenalism is if the argument succeeds. There are very many different kinds of qualia: think of how many different types of colors there are. And sounds. And emotions. Avoidance behavior will be an appropriate response to very few of these qualia types: pain, perhaps fear, perhaps some sounds (e.g., annoying high pitched sounds) etc. That is to say, the odds that avoidance behavior would be associated with appropriate qualia types as a function of pure chance are rather low. And the same can be said for any other appropriate qualia-behavior

<sup>4</sup> Sprigge (1988) has also argued that pains and pleasures have their causal powers necessarily because their causal powers are determined by their nature; so the claim, while surprising, is certainly not new.

pairing. And the odds that even a handful of appropriate pairings would obtain are very, very low (the odds of one improbable thing happening are small, but the odds that numerous improbable things would happen are extremely small). In short, given epiphenomenalism, the fact that some qualia are paired with appropriate behaviors, in the manner that they actually are, is so remarkable, so highly improbable, that one is tempted to simply conclude that epiphenomenalism is false.

Second, one might object that the argument seems to work well for certain types of qualia but not for others. It seems there is a certain fit between, say, pain or happiness and the behavior they cause, but such a fit is less obvious with other types of qualia. That is, perhaps with some types of qualia, it is not obvious what the appropriate behavior would be. This objection seems reasonable, but the argument can succeed even if the point is conceded. All the argument requires is that a fit obtains between a handful of types of qualia and behaviors, because the probability of even a handful of appropriate alignments obtaining through pure chance is very low. It seems plausible that we can find enough cases in which an appropriate alignment obtains, even if there are cases that lack such an alignment. Arguably, colors provide further examples of appropriate fits between qualia and behaviors. Think of the well-known research concerning “hot” and “cold” colors. For instance, blue qualia relax us; given the qualia, this is perhaps the behavior we would expect. It also seems that “emotional qualia” are often associated with appropriate behaviors as well; e.g., love and rage. So, it is not implausible that we can find enough appropriate alignments to allow the argument to succeed.

Third, similarly, one might object: suppose there are cases in which there is an appropriate alignment between qualia and behavior and cases in which there is not; again, maybe some qualia-behavior pairings have a high degree of fit, while others do not? If so, then maybe all the argument can show is that some types of qualia, i.e., those that play a role in an appropriate alignment, have causal powers. Perhaps, but this weaker result would still be significant. Also, there is something odd about this objection: why would some types of qualia have causal powers but others wouldn't? That is, it seems plausible to infer that all qualia have causal powers if some do, to avoid positing a strange non-uniformity between various types of qualia.

Fourth, one might object that given the appropriate alignments between qualia and behavior, it is inferred that qualia likely cause behavior. But on this view, it is still a mystery why these appropriate relations obtain. That is, even on this view, a given type of qualia might as well cause a different, inappropriate behavior. But this misses the point: qualia often cause the behaviors that they do precisely because of their subjective natures. In other words, the appropriate relationships between qualia and behavior are not astounding on this view, because the reason certain types of qualia cause the behaviors that they do is their subjective natures. It isn't as if a given qualia type could just as easily cause a different, inappropriate behavior, because the very nature of the qualia determines what behavior the qualia cause. For example, pain is unpleasant, and this is precisely why it causes some of the behaviors that it does; there is no mystery here.

Fifth, one might also object: posit a type of quale  $Q$ , a type of behavior  $B$  and a type of

brain state  $S$ . One might claim that both  $Q$  and  $B$  have a common cause, namely  $S$ . This explains why  $Q$  and  $B$  are always found together, and on this view,  $Q$  does not cause  $B$  (or perhaps anything else). That is, we have an explanation for the alignment, but qualia might be epiphenomena nevertheless.<sup>5</sup> But this response fails.  $S$  causes two disparate things,  $Q$  and  $B$ . But it is still mysterious why  $Q$  and  $B$  are appropriate for one another. That is, if qualia are epiphenomena, then  $S$  might as well cause a different type of quale,  $Q'$ , that is inappropriate for  $B$ . In short, on this view, it is still astounding that qualia and behavior are often appropriately aligned; this view does not remove the problem, but merely shifts it around. A dualist might object: dualistic epiphenomenalism claims that the physical causes qualia. So, if  $S$  causes a type of qualia  $Q$  in the actual world,  $S$  will always cause  $Q$  in the actual world; and, for that matter, the same will hold in all worlds with the same laws of nature, i.e., the causal relation is nomologically necessary. So, it is not the case that “ $S$  might as well cause a different type of quale,  $Q'$ , that is inappropriate for  $B$ .” Quite simply,  $S$  cannot cause a different type of qualia, at least in our world. But the point is that if epiphenomenalism is true, it is surprising that the laws of nature in our world are what they are; i.e., it is surprising that  $S$  causes a type of qualia that is appropriate for  $B$ . A dualist might respond that the laws of nature are contingent, so there will be many worlds in which such fits obtain and many in which they do not, so it is not terribly surprising that we are in a world where they do obtain. But if the relevant causal laws are contingent, clearly there will be many more worlds in which these fits do not obtain than worlds in which they do; so it is still highly improbable that we would be in a world in which they obtain. To put the point a different way: we have two competing explanations for the appropriate alignments that obtain between qualia and behavior in our world. These alignments obtain either (i) because qualia cause behavior, and the subjective nature of qualia determine the specific behaviors that they cause, and so these relations are appropriate, and qualia have the same causal powers in all possible worlds, or they obtain (ii) because there are worlds in which these appropriate alignments happen to obtain (even though there are many more worlds in which they do not), and we happen to be in one of the worlds that has a certain degree of “alignment.” We have two competing explanations for some state of affairs (namely, the appropriate alignments), neither of which is obviously *prima facie* absurd; but (i) makes the probability that the state of affairs would obtain 1 (indeed, it makes the state of affairs necessarily true), while (ii) makes the state of affairs highly improbable. We should opt for (i). *Ceteris paribus*, if  $P$  is true, and there are two competing explanations for  $P$ ,  $A$  and  $B$ , and  $A$  entails  $P$  while  $B$  makes it highly improbable that  $P$ , explanation  $A$  is better.<sup>6</sup>

One might claim that perhaps the common cause of  $Q$  and  $B$  is appropriate for  $Q$  and  $B$  in some sense; this appropriateness explains why  $Q$  and  $B$  are appropriate for one another, and  $Q$  might be epiphenomenal all the same. But we must be told what this common cause is. Also, we must be shown that this common cause is “appropriate” for both  $Q$  and  $B$ . This might be difficult; it seems that in a vast majority of cause and effect relationships, there is

<sup>5</sup> Robinson (2003) raises precisely this objection to James’s argument alluded to above.

<sup>6</sup> Someone might parody the argument: Joe has a yellow shirt on today. Either this is because (i) it is a necessary truth that Joe has a yellow shirt on today, or (ii) there are some worlds where Joe happens to have a yellow shirt on today (even though there are many more worlds in which he doesn’t), and this is one of them. We should endorse (i). But it’s bizarre, to say the least, to claim that Joe necessarily has a yellow shirt on. The problem with this objection is that (i), in this example, is clearly absurd; the probability that Joe necessarily has a yellow shirt on is likely zero. So the probability that (i) is true is very low. That is to say, the principle used above should not be used when one of the putative explanations for some state of affairs is obviously absurd.

no “appropriateness” component (e.g., if we heat water, the water boils, but boiling doesn’t seem to be an “appropriate companion” to heat in the same way that some behaviors seem to be appropriate companions for some qualia). Another issue concerns the nature of the “appropriate” relation, specifically, is this relation transitive? If it is not, then the putative common cause might be appropriate for  $Q$ , and it might be appropriate for  $B$ , but  $Q$  and  $B$  might not be appropriate for one another, in which case this response to my argument fails. Granted, it seems like the “appropriate” relation would be transitive, but it seems like the “similarity relation” would be transitive as well, even though it doesn’t appear to be (it appears possible that some  $A$  might be similar to some  $B$  and  $B$  might be similar to some  $C$  even though  $A$  is not similar to  $C$ ). Further, presumably this common cause will be a brain state, and it is difficult to see how a brain state can be an appropriate cause of a behavior in the same way that qualia appear to be.

Sixth, one might hold the following view: a type of qualia  $Q$  is correlated with (or is frequently found with) a belief (or doxastic state)  $D$ . For instance, pain qualia are correlated with the belief that pain qualia are subjectively unpleasant; where one has pain qualia, one also often finds the *belief* that pain qualia are unpleasant. Furthermore,  $D$  might very well be the cause of  $B$ . This leaves open the possibility that  $Q$  causes nothing, i.e., qualia are epiphenomena. But now another puzzle arises: if qualia are epiphenomena, then it seems mysterious why they are correlated with the appropriate doxastic states. For example, pain qualia are intrinsically unpleasant; they are unpleasant in and of themselves. But if pain qualia do nothing in a causal sense, then they might as well be correlated with different doxastic states; for example, pain qualia might as well be associated with the belief that pain qualia are pleasant. So, just as it was mysterious why certain types of qualia are associated with appropriate *behaviors* if qualia are epiphenomena, this view makes it mysterious why certain types of qualia are associated with appropriate *beliefs* if qualia are epiphenomena. One could modify the argument (in section one) to conclude that qualia cause certain beliefs (as opposed to behaviors), and call epiphenomenalism into question all the same.

Seventh, above, I claimed that emotional qualia are sometimes associated with appropriate behaviors. One might claim that the explanation for this is that the behavior itself is a constituent part of the emotion (see Goldie (2000), for example, who claims that emotions involve dispositions to behave in certain ways). The problem with this objection is that we still need an explanation for why certain behaviors are associated with the appropriate qualia. If the subjective natures of qualia play no role in determining behavior, then again, any behavior could have been associated with any type of qualia.

Eighth, I argued that qualia are often associated with appropriate behaviors, and this suggests that qualia cause the behaviors; for example, pain qualia cause avoidance behavior. But there are rare examples in which an agent appears to have pain qualia but does not display avoidance behavior or any other behavior typically found with pain (say, wincing). Indeed, the agents claim to feel pain, even intense pain, yet do not seem to be bothered by it. This condition has been thought to arise from prefrontal lobotomies, hypnosis, nitrous oxide, or certain drugs such as morphine (see Aydede (2013, section 5.1) for an interesting

discussion of this issue).<sup>7</sup> But if pain qualia can be present in the absence of the behaviors generally associated with pain, qualia do not cause these behaviors, and my argument fails. It seems plausible, however, that the pain qualia in question still have their normal causal powers, e.g., they still have the causal power to cause avoidance behavior or wincing; it is simply that the qualia are prevented from exercising these causal powers. So, for instance, nitrous oxide or morphine interferes with the brain in such a manner that prevents pain qualia from causing avoidance behavior or wincing, even though they still have, and always will have, the causal power to do so. There is nothing strange about an entity being prevented from exercising its causal powers. For example, a massive amount of water has the causal power to destroy a town, but perhaps it is prevented from exercising this causal power by a dam. We wouldn't claim, for instance, that the water lacks the causal power in question simply because it is prevented from exercising it.<sup>8</sup>

Ninth, one might object that the appropriate relation between pain and avoidance behavior does not hold in cases of masochism. Here are instances in which an agent has pain qualia that are not accompanied by avoidance behavior; if so, pain qualia do not cause avoidance behavior. But again, a plausible response is to claim that the pain qualia in question have the causal power to cause avoidance behavior and wincing, but the qualia are prevented from exercising these causal powers for some reason. In fact, even masochists no doubt react in normal ways to pain in some or even many cases (e.g., if someone sneaks up behind them and hits their foot with a hammer), and have limits as to how much pain they can withstand before they display the normal reactions to it; this suggests that a masochist's pain has the usual causal powers, but these causal powers are simply nullified or overridden sometimes. Perhaps the brain of a masochist simply has a different physical organization than the brain of a non-masochist, and this difference in organization prevents pain qualia from exercising their causal powers at times? Also note that we seem to view masochism as almost a quasi-logical anomaly. It strikes us as very odd that pain qualia can lack some of the causal powers they are thought to have; perhaps a more plausible explanation is that pain qualia never lack these causal powers, but are simply prevented from exercising them in some instances.

Tenth, the epiphenomenalist might attack the claim that qualia have at least some causal powers necessarily by arguing, for example, that it is conceivable that qualia have different causal powers, so it is possible that qualia have different causal powers. For instance, we can imagine a world where pain does not cause avoidance behavior, but causes people to, say, lift their right arm; so, there is a world in which this is the case. Further, if the fit intuition entails that qualia have some causal powers necessarily, it must be false, so the argument cannot even show that qualia have causal powers in the actual world. It is conceded that we can conceive of such worlds; this seems undeniable. Indeed, such scenarios are probably even easier to imagine than zombies. However, following many

---

<sup>7</sup> Philosophers might be familiar with such issues through Dennett's (1978) classic paper "Why You Can't Make a Computer that Feels Pain."

<sup>8</sup> One might wonder what more banal examples of pain imply for the argument; e.g., think of a case in which someone has a headache but takes aspirin. Such cases do not affect the argument: the aspirin will cause the pain qualia to go away, but for my argument to be falsified, we need cases in which pain qualia are present but the typical behaviors are not.

others, it is denied that conceivability is an accurate guide to possibility.<sup>9</sup> Many have claimed, for example, that we can conceive of worlds in which water is not H<sub>2</sub>O; yet if Kripke (1972) is correct, such worlds are impossible. Here is a slightly different argument: some can imagine that there are worlds in which pain, for example, causes entirely different sorts of behavior than it causes in the actual world. But some can also imagine that such worlds do not exist, i.e., it is conceivable to some that pain has its causal powers necessarily; the idea is not incoherent or logically inconsistent, for example. After all, the claim was endorsed above. But then, if conceivability entails possibility, worlds in which pain has different causal powers than it has in the actual world are both possible and impossible, a contradiction. Therefore, conceivability does not entail possibility.<sup>10</sup>

Eleventh, one might claim that although we do judge that there is a fit between qualia and behavior, this judgment is false (judgments that are about qualia, such as the one we are discussing here, are generally called “phenomenal judgments” in the literature). Perhaps we are simply hardwired in such a manner that we form this (incorrect) phenomenal judgment? However, this objection is problematic. It is strange that we would be predisposed to form an incorrect phenomenal judgment. What purpose does this judgment serve? Why is it there? Why, e.g., would we evolve to have a predisposition to form an incorrect phenomenal judgment? Why should we think the judgment is false (aren’t many, if not most, of our phenomenal judgments accurate)? The objection has no force unless we are given a reason to reject the judgment. Perhaps we judge that there is a fit between behavior and qualia simply because some types of qualia are generally associated with certain behaviors, and we think there must be some reason for this association? But then why does it seem that the fit is stronger in some cases than in others? And it is clear that we see cases of constant conjunction all of the time, yet don’t judge that the relation between the relevant entities is *appropriate*. We might judge that there is a causal relation at work, but not necessarily an appropriate relation. So why would we do so in the case of qualia and behavior?

Twelfth, one might object that a necessary truth is inferred from a probabilistic argument, and it is difficult to see how an argument from probability can establish a necessary truth. But this objection misconstrues the structure of the overall argument. The fit intuition, when combined with certain other claims (e.g., the observation that the fit is not a result of pure chance), entails that epiphenomenalism is very improbable. Moreover, the fit intuition entails that these causal powers are metaphysically necessary. A necessary truth is not inferred from the probabilistic argument *per se*; rather, it is the fit intuition that suggests that these causal powers are metaphysically necessary. Basically, the fit intuition suggests both that qualia have causal powers in the first place, and it also suggests that these causal powers are metaphysically necessary.

<sup>9</sup> See, for example, Hill and McLaughlin 1999, 446; Hill 1997; Loar 1999; and Block and Stalnaker 1999; of course, often this literature is an attempt to undermine Chalmers’s (1996) zombie argument. See also Kirk (2012, section 5.1) for a discussion of this issue.

<sup>10</sup> See also Winstanley (2007); Winstanley argues that, at least on Chalmers’s account of conceivability and possibility, the claim that conceivability entails possibility implies a contradiction (because we can conceive that a necessary being exists and we can conceive that such a being does not exist). Winstanley (2007) argues that this problem even applies to Chalmers’s more recent, and more sophisticated, account of the relationship between conceivability and possibility (see, e.g., Chalmers 2002). Another issue is this: presumably, our ability to conceive of certain situations is a cognitive ability that is, for example, carried out by some part of the brain. It is difficult to see why we should think this ability is perfect, or this area of the brain is infallible etc.

In sum, there are numerous possible objections to the argument: some deny that there are appropriate relations between qualia and behavior at all; some point out that appropriate relations obtain only sometimes; some grant that the appropriate relations obtain but try to find some other explanation for them than the one I offer; still other objections try different strategies. However, there are responses to these objections.

#### 4. Dualism

Thus far, I've argued that (i) qualia have causal powers, i.e., epiphenomenalism is false, (ii) different types of qualia have different causal powers, and (iii) qualia have at least some of their causal powers necessarily. These claims, if true, have some implications for dualism, which I now discuss.

First, of course, dualistic epiphenomenalism is a prominent form of dualism; indeed, it has been perhaps the most common form of dualism throughout dualism's history. But if epiphenomenalism is likely false, dualistic epiphenomenalism is likely false. It seems the only viable form of dualism is interactionism; note that the argument in section one also shows that parallelism is likely false; in fact, the argument applies to any view that denies that the subjective natures of qualia have causal powers. Second, any worry that dualism might be committed to epiphenomenalism, e.g., the concern over the causal closure of the physical, or more traditionally, the concern over the conservation of energy, becomes even more pressing; if epiphenomenalism is probably false, then dualists cannot simply endorse epiphenomenalism to avoid such worries.

The dualist might object: it seems that your argument causes serious problems for certain forms of physicalism too, e.g., any form of physicalism that denies that the subjective natures of qualia have causal powers. And, just as it makes, say, worries over causal closure more pressing, doesn't it also make worries over Kim's (2000; 2005) exclusion argument more pressing? I grant this objection: any view that denies that the subjective natures of qualia have causal powers, whether the view is a form of dualism or a form of physicalism, is likely false. To elaborate, a dualist might object that the argument calls functionalism, the view endorsed by many contemporary physicalists, into question. Functionalism, at least in its most commonplace, basic form, claims that what determines the nature of, say pain, are the inputs that lead to pain (say, someone stepping on your foot) and the outputs that pain produces (including certain behaviors). But if the nature of pain is determined by what it causes, at least in part, and not vice versa, then it seems that any type of qualia could have been associated with pain (e.g., pain might as well feel like happiness, it wouldn't make any difference since the way that pain feels plays no role in determining what it causes). I do not know if this objection against functionalism succeeds or not, but assuming that it does, it is granted as well. If functionalism must deny that the subjective nature of qualia does not help determine what behaviors the qualia cause, then it is probably false.

Finally, there is, of course, much debate concerning the modal strength of the laws of nature: some hold they are metaphysically necessary (e.g., Wilson 2005; Shoemaker 1980, 1998; Swoyer 1982; Bird 2005) while many others hold they are metaphysically contingent (or are only nomologically necessary). A commonplace assumption in this debate is

that all causal laws are metaphysically necessary or all causal laws are metaphysically contingent; it is not the case that some laws are contingent while others are necessary. I do not know if this assumption is true or not. However, there are some reasons for thinking that it is true. One reason is that generally when an argument seeks to show that laws are metaphysically necessary or contingent, the argument is such that it applies to *all* instances of causation. So, e.g., one might think the laws of nature are metaphysically contingent because one endorses Hume's dictum, the claim that there cannot be necessary connections between distinct entities; but if Hume's dictum is true, *all* causal laws are metaphysically contingent. Therefore, if some causal laws are contingent while others are necessary, practically everyone that has participated in the debate, and practically all of the arguments that have been formulated for either side, are mistaken. Also, perhaps the modal strength of a relation is such an important – central – aspect of the relation that it is an essential aspect of the relation? Moreover, one could argue that the burden of proof is on those who argue that causal relations can have different modal strengths; if one is going to posit a novel metaphysical asymmetry, plausibly, they must justify it. At any rate, *if* this assumption is true, i.e., *if* it is the case that all causal laws have the same modal strength, then given that at least some of the causal laws that involve qualia are metaphysically necessary (section two), then all causal laws are metaphysically necessary.

And given that interactionism claims that the physical causes qualia, the same brain state types will cause the same qualia types in all metaphysically possible worlds; but then qualia will strongly supervene on the physical. But note that some dualists have sometimes associated dualism with the claim that qualia only nomologically supervene on the physical (whereas physicalism is associated with the claim that qualia strongly supervene on the physical); Chalmers (1996) is a notable example. But if qualia strongly supervene on the physical, then the dualist will be unable to characterize dualism as the view that qualia only nomologically supervene. Indeed, we might need a new way to characterize dualism and physicalism alike if metaphysical and nomological supervenience collapse into one another. Moreover, some of the best arguments for dualism argue that qualia only nomologically supervene on the physical. But if qualia strongly supervene on the physical, then by definition, physical duplicates cannot have inverted qualia and phenomenal zombies will be impossible. In other words, dualism will be unable to appeal to much of its current justification (see Byrne (2010) for more on inverted qualia arguments for dualism and Chalmers (1996) for the zombie argument). Of course, dualism still might be true. It is simply that now we have less reason to believe it.<sup>11</sup>

---

<sup>11</sup> I would like to thank anonymous referees for their valuable suggestions.

## REFERENCES

- Aydede, M. 2013. Pain. *The Stanford Encyclopedia of Philosophy (Spring 2013 Edition)*. Edited by Edward N. Zalta. Accessed February 17, 2014. <http://plato.stanford.edu/archives/spr2013/entries/pain/>.
- Bird, A. 2005. The Dispositionalist Conception of Laws. *Foundations of Science* 10: 353-370.
- Block, N. and R. Stalnaker. 1999. Conceptual Analysis, Dualism, and the Explanatory Gap. *Philosophical Review* 108: 1-46.
- Byrne, A. 2010. Inverted Qualia. *The Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*. Edited by Edward N. Zalta. Accessed February 17, 2014. <http://plato.stanford.edu/archives/spr2010/entries/qualia-inverted/>.
- Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York and Oxford: Oxford University Press.
- Chalmers, D. J. 2002. Does conceivability entail possibility? In *Conceivability and Possibility*, ed. T. S. Gendler and J. Hawthorne, 145-200. Clarendon Press: Oxford.
- Dennett, D. C. 1978. Why You Can't Make a Computer that Feels Pain. In *Brainstorms*, 190-229. Cambridge, MASS: MIT Press.
- Dennett, D. C. 2005. *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MASS: MIT Press.
- Goldie, P. 2000. *The Emotions: A Philosophical Exploration*. Oxford: Oxford University Press.
- Harman, G. 1965. The Inference to the Best Explanation. *Philosophical Review* 74: 88-95.
- Heil, J. 1998. *Philosophy of Mind: a contemporary introduction*. London and New York: Routledge.
- Hill, C. S. 1997. Imaginability, Conceivability, Possibility and the Mind-Body Problem. *Philosophical Studies* 87: 61-85.
- Hill, C. S. and B.P. McLaughlin. 1999. There are Fewer Things in Reality Than Are Dreamt of in Chalmers's Philosophy. *Philosophy and Phenomenological Research* 59: 446-454.
- James, W. 1879. *Are We Automata?* *Mind* 4:1-22.
- James, W. 1890. *The Principles of Psychology*. H. Holt.
- Kim, J. 2000. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MASS: MIT Press/Bradford Books.
- Kim, J. 2005. *Physicalism, or Something Near Enough*. Princeton and Oxford: Princeton University Press.
- Kirk, R. 2006. Zombies. *The Stanford Encyclopedia of Philosophy (Summer 2012 Edition)*. Edited by Edward N. Zalta. Accessed February 17, 2014. <http://plato.stanford.edu/archives/sum2012/entries/zombies/>.
- Kripke, S. 1972. Naming and Necessity. In *Semantics of Natural Language*, ed. D. Davidson and G. Harman, 253-355. Dordrecht, Holland: Reidel.
- Loar, B. 1999. David Chalmers's *The Conscious Mind*. *Philosophy and Phenomenological Research* 59: 465-472.
- Robinson, W. 2003. Epiphenomenalism. *The Stanford Encyclopedia of Philosophy*

- (*Spring 2003 Edition*). Edited by Edward N. Zalta. Accessed March 7th, 2006. <http://plato.stanford.edu/archives/spr2003/entries/epiphenomenalism/>.
- Shoemaker, S. 1980. Causality and Properties. In *Time and Cause*, ed. P. van Inwagen. Dordrecht: D. Reidel Publishing Company.
- Shoemaker, S. 1998. *Causal and Metaphysical Necessity*. *Pacific Philosophical Quarterly* 79: 59-77.
- Sprigge, T. L. S. 1988. *The Rational Foundations of Ethics*. London: Routledge and Kegan Paul.
- Swoyer, C. 1982. *The Nature of Natural Laws*. *Australasian Journal of Philosophy* 60: 203-223.
- Tye, M. 2013. Qualia. *The Stanford Encyclopedia of Philosophy (Summer 2008 Edition)*. Edited by Edward. N. Zalta. Accessed February 17, 2014. <http://plato.stanford.edu/archives/fal2013/entries/qualia/>.
- Wilson, J. 2005. Supervenience-based Formulations of Physicalism. *Nous* XXXIX, 3: 426-459.
- Winstanley, P. 2007. Conceivability, Possibility, and Two Dimensional Semantics. *Percepti* 1: 18-31.
- 

Received: March 24, 2012

Accepted: January 3, 2013

Carroll College  
1803 Poplar Street  
Helena MT, 59601 USA  
[jlm3am@virginia.edu](mailto:jlm3am@virginia.edu)