

Marcus Damm Strøm-Hansen:

Epiphenomenalism and Knowledge – Defending Robinson’s Reply to the Self-Stultification Objection

RESUMÉ

Ifølge epifænomenalisme har mentale hændelser ingen kausal indflydelse på fysiske hændelser. En af de stærkeste indvendinger mod dette synspunkt er den såkaldte indvending om selvundergravelse, som forsøger at vise, at epifænomenalisme er inkompatibel med vores viden om egne oplevelser. William Robinson (2006) argumenterer for, at epifænomenalisme kan undslippe beskyldningerne om selvundergravelse ved at appellere til den underliggende årsag, som vores mentale hændelser og vores ytringer om mentale hændelser har til fælles. I denne artikel forsvarer jeg Robinsons position mod en række indvendinger, som Dwayne Moore (2012) rejser. Jeg konkluderer at Robinsons løsning på indvendingen om selvundergravelse består.

ABSTRACT

Epiphenomenalism is the view that mental events have no causal impact on physical events. One of the most potent objections to this view is the self-stultification objection, which aims to show that epiphenomenalism is incompatible with knowledge about our own experience. William Robinson (2006) argues that epiphenomenalism can escape charges of self-stultification by appealing to the common underlying cause between mental events and our reports of mental events. In this paper, I defend Robinson’s proposal against several objections raised by Dwayne Moore (2012). I conclude that Moore’s arguments fail to undermine Robinson’s solution to the self-stultification objection.

EMNEORD

Bevidsthed, epifænomenalisme, viden

KEYWORDS

Mind, consciousness, epiphenomenalism, knowledge

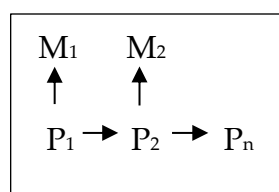
1. Introduction

Epiphenomenalism is the view that mental events have no causal impact on physical events. While most people—philosophers and laymen alike—regard this idea as highly implausible, even absurd, I think there are strong reasons to accept it. Still, it faces obstacles, and the self-stultification objection is among the most potent. This objection aims to show that epiphenomenalism is incompatible with knowledge about our own experience. Robinson (2006) argues that epiphenomenalism can escape charges of self-stultification by appealing to the common underlying cause between mental events and our reports of mental events. In this paper, I defend Robinson’s proposal against several objections raised by Moore (2012).

The paper has the following structure: In section 2, I define epiphenomenalism and show that it follows from two relatively plausible claims: irreducibility of qualia (2.1) and causal closure (2.2). In section 3, I present the self-stultification objection. Then, in section 4, I reconstruct and discuss Robinson’s proposed solution to the self-stultification objection. Finally, in section 5, I consider three objections raised by Moore: the Irrelevance Objection (5.1), the Problem of Psychophysical Laws (5.2), and the Epistemic Luck Objection (5.3). I argue that all three objections fail to undermine Robinson’s solution to the self-stultification objection.

2. Motivating Epiphenomenalism

In its most general form, *epiphenomenalism* holds (1) that mental events are caused by physical events and (2) that mental events have no causal impact on physical events. This gives us the following picture of the relationship between physical and mental events:



The physical event P_1 causes another physical event P_2 *and* a mental event M_1 . By contrast, the two mental events M_1 and M_2 do not cause any physical events.¹ In a sense, these epiphenomenal mental events are something “extra,” like the steam whistle on a locomotive or a shadow tracing the steps of its master. They contribute nothing to the causal affairs of the physical world (Robinson 2019).

Recent authors define epiphenomenalism more narrowly as a view about *qualitative events* rather than mental events per se (Robinson 2006). This requires some explaining: Qualitative events are mental events that have *phenomenal properties*, meaning there is something it is like to have them. For instance, my experience of pain after being sunburned is a qualitative event because there is something it is like for me to have this pain-experience; it has the phenomenal property of being painful. There are many other ways of expressing roughly the same idea: Jackson (2002) uses the term “qualia” to denote the qualitative properties of mental events, Block (2002) talks of mental states being “phenomenally conscious,” and Chalmers (2004) seems to prefer the simple “consciousness.”

Qualitative events are commonly distinguished from another type of mental event, namely, *propositional attitudes*, such as beliefs and desires. This distinction is often held to be metaphysically important. It is argued that propositional attitudes can be analyzed in functional terms and reduced to physical events, whereas qualitative events cannot—they are irreducible.² On this view, propositional attitudes can be fitted into the causal workings of the physical world without much fuss. But the causal status of qualitative events is more mysterious, hence the motivation for epiphenomenalism about qualitative events. I will return to these points in the next section.

Now if we replace “mental events” with “qualitative events” in our definition of epiphenomenalism, we get what Moore (2012, 628) calls *qualia epiphenomenalism*: Qualitative events are caused by physical events but have no

¹ Note that epiphenomenalism thus defined does not rule out that mental events can have causal effects on other mental events. So there could be a causal chain of mental events as long as none of these mental events have physical effects.

² See, for example, Chalmers (2004).

causal efficacy on the physical world. In the rest of this paper, I will use “epiphenomenalism” as shorthand for “qualia epiphenomenalism.”

With these definitions in place, we can move on to assess the plausibility of epiphenomenalism. At first, it may seem like a counterintuitive, even patently false, view. After all, is my screaming not caused by my experience of pain? Is my decision to wear sunscreen not caused by my vivid memory of a nasty sunburn? It seems strange that the answer to these questions should be no. However, it turns out that epiphenomenalism is a consequence of two plausible (though controversial) claims: *irreducibility of qualia* and *causal closure*. In the following, I will briefly present these claims and show how, when combined, they motivate epiphenomenalism.

2.1 Irreducibility of Qualia

According to the thesis which I will call *irreducibility of qualia*, some properties of mental events, namely, phenomenal properties, or qualia, are not reducible to any physical properties. On this view, the sharp smell I experience when putting on hand sanitizer cannot be reduced to the firing of neurons in my brain or to any functional state – it is something over and above the physical.

Irreducibility of qualia is certainly controversial. However, some powerful arguments support it. Let me briefly sketch one of them: *the knowledge argument* put forward by self-professed “qualia freak” Frank Jackson (2002, 275).³ Jackson asks us to imagine a brilliant scientist, Mary, who knows all the *physical information* there is to know about color and vision, about the wavelengths of light, and the workings of the nervous system. Unfortunately, Mary has been locked in a completely colorless room her whole life, so she has never seen a colored thing. She has never seen a red tomato or the deep blue ocean. Now, what happens when Mary is released from her colorless room and experiences color for the first time? Does she learn anything? She already possesses the relevant physical information, so if she gains new information, it must be non-physical.

³ Jackson has since changed his mind and is now a professed physicalist.

According to Jackson (2002, 275), Mary does learn something. She learns *what it is like* to experience color. Therefore, physical information is not the only information we can have; there is also what we might call *phenomenal information* which concerns the felt properties of experience. In other words, there are properties—qualia—that are not captured by the physicalist picture of the world, and so the thesis of irreducibility of qualia is true.

The knowledge argument has sparked a lively debate which has given rise to a myriad of objections and counterarguments.⁴ To further complicate matters, the knowledge argument is only one path to irreducibility of qualia. There are also *modal arguments* that reason from the possibility of zombies—creatures who are exact physical copies of us but who lack qualia—to the conclusion that qualia are non-physical (Jackson 2002, 275); and the reductionist also has an arsenal of objections against this style of argument. My aim here is not to settle the debate. However, I have demonstrated one way that the epiphenomenalist might motivate the thesis of irreducibility of qualia.

2.2 Causal Closure

By itself, the thesis of irreducibility of qualia does not entail epiphenomenalism. To complete the epiphenomenalist picture, we must also assert the principle of *causal closure of the physical*, which states that any physical event that has a sufficient cause has a sufficient *physical* cause. Support for this claim can be found in empirical science. We have a detailed understanding of neurophysiology and the interactions between neurons and other cells in the body; human and animal behavior can—in theory at least—be wholly explained in physical terms. For instance, my utterance “I am in pain” is caused by the muscles of my mouth and abdomen, which are activated by neurons firing in my brain, which, in turn, receive their cue from sensory neurons on the surface of my sunburned skin. So, my utterance has a sufficient physical cause, and so does every other physical event (that has a sufficient cause).

Now, causal closure as defined above does not rule out mental to physical causation. It could be that behavior has mental causes in addition to sufficient physical causes. To rule out this possibility, we need the further premise that

⁴ See Nida-Rümelin and O’Conaill (2019) for an overview of the different arguments.

our behavior is not *causally overdetermined*. Is this a plausible assumption? First, we should note that some physical phenomena *are* causally overdetermined. For instance, if I am swimming in the sea while it rains, my being wet is overdetermined as both the seawater and the rain are sufficient causes. However, as Robinson (2019) argues, causal overdetermination is not very plausible in the case of mental to physical causation for the following reasons:

First, it goes against the epistemic value of simplicity to postulate unnecessary causes.⁵ One might think that this kind of methodological principle is too weak to overrule the powerful intuition that our mental events really do have an impact on our behavior. But this intuition has already been denied once we accept the principle of causal closure. If any physical event (that has a sufficient cause) has a sufficient physical cause, then mental causes must be superfluous; it makes no difference whether they are there or not. In other words, causal overdetermination fails to support the main anti-epiphenomenalist intuition that mental events *make a difference* to our behavior. Add to this the epistemological predicament that, insofar as mental causes make no difference to behavior, they would not figure in our scientific theories, and so we could not know if they were there (Robinson 2019).

None of these arguments are conclusive. In principle, it is still possible to hold that behavior is causally overdetermined, that there are both sufficient physical causes *and* mental causes. However, for the reasons sketched above, it does not seem like a particularly attractive view. Thus, given that our behavior is explainable fully in terms of synapses, muscle contractions, and other physiological terms and given that overdetermination by mental causes is quite implausible, we ought to accept the conclusion that qualitative events have no causal impact on the physical world.

This completes the epiphenomenalist picture – almost. We need to add the final assumption that qualitative events are caused by physical events. It is uncontroversial that brain processes are correlated with consciousness; most obviously, damage to the brain is often followed by changes in experience. And

⁵ The physicalist could accuse the dualist of having already violated simplicity in asserting the existence of irreducible qualia. However, the dualist thinks there are strong reasons to think that they *do* exist even if it complicates our picture of the world.

even though it remains a mystery how causation between neural events and qualitative events could occur, the alternative is much less plausible; *parallelism*—in addition to facing many of the same objections as epiphenomenalism—requires us to think that the physical and the mental are synchronized in some cosmic harmony. So, it seems reasonable to suppose that qualitative events are caused by physical events.

To sum up, epiphenomenalism follows from some relatively plausible claims: (a) qualitative events are irreducible to physical events, (b) they are caused by physical events, (c) but they have no causal efficacy themselves. In the rest of the paper, I will discuss a major objection raised against this view.

3. The Self-Stultification Objection

I have already alluded to the fact that epiphenomenalism leads to some counterintuitive results. According to the epiphenomenalist, my actions are not caused by my qualitative events. This means that my thirst does not cause my getting a drink and that my pain does not cause my saying “ouch.” Rather, my experience of thirst and pain merely coincide with the physical causes underlying my behavior. This is certainly a strange conclusion, but it does not undermine epiphenomenalism—it is a reasonable bullet to bite. The real danger for epiphenomenalism is expressed in the *self-stultification objection*, which says that epiphenomenalism is incompatible with knowledge about our own qualitative events. This leads to the devastating conclusion that if epiphenomenalism were true, it could not be known to be true (Robinson 2006, 88).

Let us take a closer look at this argument. It concerns reports about qualitative events; such reports can, according to Chalmers (2004, 176), be placed in three categories. *First-order* reports concern the objects of qualitative events rather than the events themselves or their properties. For instance, if I report that “the sky is a beautiful shade of blue today,” I am making a claim about the sky and not about my experience. *Second-order* reports are about the qualitative events themselves. I might, for instance, notice that I am having a particular blue sensation, or I might report that “I am experiencing pain.” *Third-order* reports are about qualitative events as a category. An example could be “Qualitative events are strange phenomena.”

The self-stultification objection, as expressed by Robinson (2006, 88), emerges when considering second-order reports. According to epiphenomenalism, my experience of being in pain does not causally contribute to my report that “I am in pain.” This is because the former is a qualitative event and the latter is a physical event with entirely physical causes. But if my report that “I am in pain” is not in any way caused by my experience of pain, then how can the report be justified? How can it express knowledge? In more general terms, the self-stultification objection has the following structure:

(S1) A report that I have an *F* qualitative event (where *F* is any quale) can express knowledge only if the occurrence it reports causally contributes to the making of the report.

(S2) Qualitative events, according to epiphenomenalism, do not causally contribute to reports that one has them. So,

(S3) A report that I have an *F* qualitative event cannot express knowledge, according to epiphenomenalism. (Robinson 2006, 88)

In addition to being counterintuitive, the conclusion (S3) undermines some key claims in support of epiphenomenalism. For example, as we saw in section 2.2, epiphenomenalists reason from claims that we experience certain qualitative events (e.g., pain) following certain physical changes (e.g., bodily damage) to the conclusion that physical events cause qualitative events. If we cannot trust reports about the occurrence of qualitative events, then we cannot motivate this key epiphenomenalist premise. Perhaps more fundamentally, if we cannot trust our reports about qualitative events, then we have no justification for asserting the irreducibility of qualia; Mary could not know that she gained new information upon leaving her black-and-white room. So, the thrust of the self-stultification objection is that, according to their own position, epiphenomenalists are not justified in making the kinds of claims they are in fact making.

Clearly, the epiphenomenalist must find some way of responding to the self-stultification objection. And since the argument is valid, the epiphenomenalist must dispute the truth of at least one of the premises. Let us see how this might be done.

4. Robinson's Solution: A Common Underlying Cause

Robinson (2006) defends epiphenomenalism against the self-stultification objection. Specifically, he argues that (P1), the premise that a report about a qualitative event must be caused by the qualitative event to express knowledge, is false. In the following, I will lay out and assess Robinson's argument before turning to a series of counterarguments raised by Moore (2012). I will argue that Robinson's position can be defended and that epiphenomenalism remains a plausible view about the relationship between mental and physical events.⁶

The premise (P1) that a report about a qualitative event must be caused by the qualitative event to express knowledge seems plausible at first. After all, knowledge requires justification, and surely my judgment that "I am experiencing pain" must be (at least partially) caused by my pain experience if it is to be justified. How else could it be justified? This is the crux of the self-stultification objection. The second premise (P2) that qualitative events do not causally contribute to reports about them follows trivially from the definition of epiphenomenalism, which holds that qualitative events do not have physical causes. So (P2) cannot be disputed without abandoning the position altogether. It seems, then, that the epiphenomenalist must reject (P1).

This is exactly what Robinson (2006, 89) does. He claims that (P1) only *seems* plausible because it describes a central aspect of *perceptual* knowledge. My judgment that "there is a penguin in the bathtub" is justified only if the presence of a penguin in the bathtub has causally contributed to my judgment—for example, because I have seen it with my own eyes. But just because causation plays a role in perceptual knowledge, Robinson argues, it does not follow that it is necessary for knowledge of qualitative events. So, we cannot just import the causal-perceptual model to our thinking about knowledge of our own minds.

Robinson (2006, 89) goes on to argue that the causal model, when applied to knowledge of mental events in general, or qualitative events in particular, becomes self-undermining. His argument is a classic regress argument:

⁶ For an alternative response to the self-stultification objection, see Chalmers (2004, 197), who argues that it is our *direct acquaintance* with experience—and not a causal relationship—that justifies our judgments about it.

According to the causal model, if I am to have knowledge of a mental event M_1 , then M_1 must causally contribute to another mental event M_2 , but then my knowledge of M_2 requires that it causally contributes to yet another mental event M_3 , and so on *ad infinitum*. Thus, an infinite regress would occur, and I could never have knowledge of my mental events. Therefore, the causal model is self-undermining.

I do not think this argument works. First, nothing prevents us from stopping a regress like this. For instance, we could employ a strategy analogous to higher-order theories of consciousness, that is, we could say that some qualitative events are objects of knowledge in virtue of their causal contribution to other mental events, whereas these secondary mental events are not themselves objects of knowledge.⁷ This should be perfectly fine; we do not have to suppose that *all* mental events are objects of knowledge.

However, this kind of reasoning should not even be necessary since the argument is based on a confused supposition; no regress is initiated in the first place. To see this, consider the following causal story. The qualitative event that includes my pain experience causally contributes to some number of other mental events, including a belief that I am in pain. This belief then contributes to pain behavior as well as the report that I am in pain. This report expresses knowledge about the qualitative event since the qualitative event causally contributed to it. And no regress enters the picture.

Clearly, then, the causal model is not self-defeating as Robinson (2006) claims, and so (P1) cannot be outright rejected. However, this is not detrimental to Robinson's argument. He is not required to show that the causal model fails on its own terms. To reject (P1), all he has to do is give a compelling alternative explanation of how reports about qualitative events can express knowledge – one that does not require qualitative events to contribute causally to our reports about them.

Robinson (2006, 90) provides such an explanation: Reports about qualitative events can express knowledge because “[w]e would not normally report that we had an *F* qualitative event unless the brain events that cause an *F* qualitative

⁷ See, for example, Rosenthal (1997).

event were occurring” (Robinson 2006, 90). In other words, the neural event P_1 , which causes my report that I have a qualitative event Q_1 , also causes the qualitative event Q_1 . This ensures that if I report having Q_1 , then I most likely have Q_1 . So, a reliable correlation exists between the report about the qualitative event and the qualitative event itself, and according to Robinson, this suffices for knowledge.

Let me clarify a few points here: (1) Robinson does not purport to give a comprehensive account of first-person knowledge. There is more to such an account. His position is better understood as a denial of the claim that causation is a necessary condition for first-person knowledge. (2) The claim that all reports (about qualitative events) have corresponding qualitative events does not entail that all qualitative events have corresponding reports. Having a qualitative event without making a report about it is certainly possible. (3) There is a separate question of what the *referent* of a report is—the neural event or the qualitative event? It may seem that Robinson commits himself to the strange claim that the report “I am in pain” refers to a neural event rather than the experience of pain. However, it seems to me that, even if we deny a causal relationship between the qualitative event and the report, nothing stops us from holding that the report still *refers* to the qualitative event.

Now Robinson is in a position to reject (P1). By presenting his alternative explanation of how we can have knowledge of our qualitative events, he can demonstrate that causation is not a necessary condition as (P1) claims. If he is right, then the self-stultification objection fails; epiphenomenalism *is* compatible with knowledge of qualitative events, and so it is not self-undermining.

5. Three Objections and My Replies

Moore (2012) raises three objections against Robinson’s *common underlying cause account* of how we come to have knowledge of qualitative events. First, he argues that the apparent irrelevance of qualitative events for our reporting about them poses a problem for Robinson’s view. Second, he argues that since we cannot know that the requisite psychophysical laws hold, we cannot have knowledge of our qualitative events. Third, he argues that the epistemic mechanism proposed by Robinson is unreliable because it is a consequence of

epistemic luck. I will argue that all these objections fail to undermine Robinson's view.

5.1 The Irrelevance Objection

We have seen that, on Robinson's view, a report about a qualitative event is not caused by the qualitative event itself. Rather, the report and the qualitative event share a common underlying cause: a neural event. Now, Moore's first objection is that, contrary to what Robinson claims, such a *common cause relation* is too indirect to give rise to knowledge. Specifically, the problem is that the qualitative event is *irrelevant* to the occurrence of the report. Moore presents an illustration to massage our intuitions:

A twig is floating down a stream, and it crashes into a rock in the middle of the stream, which knocks two grains of sand off the rock. One grain of sand floats to the left, the other to the right. At precisely this place, the stream forks in two, leaving one grain of sand flowing down the left stream, and the other floating down the right stream. As it turns out, the grain of sand flowing down the left stream hits the bank and comes to a stop. The grain of sand flowing down the right stream, however, floats on for 1 km before getting caught by a protruding log. (Moore 2012, 632–633)

Moore (2012, 633) makes a couple of observations: (a) The fact that the first grain hits the bank is completely irrelevant to the fact that the second grain gets caught in a log. (b) These two events share a common cause, namely, the twig hitting the rock upstream. From these two observations, he concludes that a common cause does not ensure that one effect is relevant to another effect.

Presumably, Moore wants to apply this conclusion to qualitative events; the fact that a qualitative event and the report about it have the same physical cause does not entail that the qualitative event is relevant to the occurrence of the report. And presumably, he holds the unspoken assumption that if the qualitative event is irrelevant to the occurrence of the report, then the report cannot express knowledge. This allows him to conclude that reports about qualitative events cannot express knowledge even if they share a common cause.

Initially, Moore's illustration seems to prove his point rather well—there is no relevant connection between the two grains of sand. However, on closer inspection, it becomes obvious that this is not an adequate analogy to the case of qualitative events. The illustration describes a random occurrence, but on a plausible interpretation of epiphenomenalism, a *nomological* relationship exists between physical events and qualitative events. Neural event P_1 is a necessary and sufficient condition for Q_1 and a necessary condition for the report about Q_1 .⁸ This means that if the report about Q_1 is made, then the physical event P_1 must have caused it, and since P_1 always causes Q_1 , Q_1 must also have occurred. Moore (2012, 633) acknowledges this objection and proposes that we alter the example to accommodate it. We should then imagine that the first grain of sand gets stuck on the riverbank *only if* the other grain of sand gets stuck on the log.

It seems to me that the illustration loses all its intuitive punch when altered in this way. If there really were a lawlike relationship between two grains of sand getting stuck in particular places, then it might be quite plausible to say that the first grain of sand is relevant to the second grain of sand. But Moore insists that the two grains of sand are not related in the relevant way: "The second grain of sand's hitting a log is dependent upon everything within its own causal stream [...] but fails to depend on or be closely related to, or relevant to, anything that transpires in the other causal stream" (Moore 2012, 633). This quote, I think, reveals why the objection misses the mark. Moore seems to be assuming that a *causal* link must exist between two events for one to be relevant to the other. However, his sand analogy does not make this obvious. Worse yet, this assumption is precisely what Robinson disputes in the context of qualitative events. He argues that a causal link is not necessary; a *common underlying cause* is enough.

In the end, Moore's first objection fails. The grain of sand illustration is not relevantly similar to the case of qualitative events since it does not involve a nomological relationship. And if a nomological constraint is built into the illustration, then it no longer probes our intuitions in the right way. Finally, by

⁸ P_1 is only necessary and not sufficient for the report because – presumably – other factors further down the causal chain will influence whether a report is made.

insisting that a causal relationship is necessary for relevancy, Moore begs the question against Robinson.

5.2 *The Problem of Psychophysical Laws*

As previously mentioned, epiphenomenalists generally assert that a nomological relationship exists between physical events and qualitative events. They assert that the neural event P_1 *always* causes the qualitative event Q_1 or, in other words, that the *psychophysical law* $P_1 \rightarrow Q_1$ holds. Robinson's common underlying cause account depends on the existence of such psychophysical laws because without them, we would not be justified in asserting that our reports about qualitative events actually correlate with the occurrence of the qualitative events. Moore's second objection questions our knowledge of psychophysical laws.

According to Moore (2012, 635), Robinson's view runs into problems because it supposes that psychophysical laws are *nomologically necessary* but not *metaphysically necessary*. In other words, Robinson's view is that, while $P_1 \rightarrow Q_1$ does in fact hold, it *could have been the case* that some other psychophysical law, say, $P_1 \rightarrow Q_2$ applied instead – just as it could have been the case that Donald Trump won the 2020 presidential election or that the dinosaurs were not wiped out 65 million years ago. This same idea can be expressed through the notion of possible worlds: $P_1 \rightarrow Q_1$ holds in the *actual* world, but there are *possible* worlds in which $P_1 \rightarrow Q_1$ does not hold.

Moore is likely correct in his interpretation of Robinson's view. It certainly seems metaphysically possible, on an epiphenomenalist view, that P_1 could have caused Q_2 instead of Q_1 or that we could have been zombies without any qualitative events. But why should this merely metaphysical possibility be a problem for Robinson?

Moore (2012, 635) argues that if it is metaphysically possible that $P_1 \rightarrow Q_1$ does not hold, then it is metaphysically possible for me to make a report about a qualitative event without this event occurring. This is true. But Moore goes on to claim that if it is metaphysically possible that I could make a report about an absent qualitative event, then it calls into question my knowledge of qualitative events. This conclusion does not follow. There is no reason why a mere metaphysical possibility should undermine our knowledge in the actual world.

After all, in the actual world, the psychophysical law $P_1 \rightarrow Q_1$ *does* hold, so we can safely rely on it for our knowledge about qualitative events.

Moore (2012, 635) anticipates this reply and counters with a question: *How do you know which world you are in?* How do you know that you are in a world where $P_1 \rightarrow Q_1$ holds and not in one of the many possible worlds where it does not hold? Moore demands evidence that the psychophysical law holds in the actual world.

Such evidence is difficult to come by. As Moore (2012, 636) rightly points out, it cannot arise out of introspection alone since we cannot introspect our neural activity, and so we cannot correlate neural events with qualitative events by this method. Furthermore, Moore argues, even if we could scan our neural activity in real time while introspecting our qualitative events, this would only allow us to identify *instances* of neural events and qualitative events being correlated. And this does not permit us to assert the existence of a psychophysical *law*. So, it seems that we cannot know that the requisite psychophysical laws hold in the actual world.

I think the appropriate response is something like this: Moore is right to assert that we cannot be *certain* that $P_1 \rightarrow Q_1$ holds in the actual world. We cannot be certain that our reports about qualitative events are reliably correlated with the qualitative events occurring. However, Moore is effectively raising a radically skeptical concern, and in doing so, he is demanding too much of Robinson's account. We do not have to be certain that the psychophysical law holds to rely on it for knowledge, just as we do not have to be certain that we are not systematically deceived by a Cartesian demon to trust our perceptions. Robinson is no more obligated to solve the problem of skepticism than someone defending the reliability of our perceptual apparatus.

Insofar as knowledge requires that our beliefs be generated by a reliable mechanism and insofar as a reliable correlation exists between neural events on the one hand and qualitative events and our reports about them on the other, then we *can* have knowledge of our qualitative events. Of course, there is always the possibility that this correlation is unreliable, that we are living in a zombie world, or that we are being deceived by an evil demon. Our knowledge is fallible, but it is still knowledge.

5.3 The Epistemic Luck Objection

Moore (2012, 638) begins his third and final objection by granting, for the sake of argument, that the requisite psychophysical laws can be put in place such that our reports about qualitative events really do correlate with the right qualitative events. Even so, he argues, the epistemic mechanism, that is, the existence of a common cause underlying both the report and the reported event, is unreliable because it is a product of epistemic luck. He concludes that the truth of our reports is accidental and, consequently, that they cannot express knowledge.

The problem of *epistemic luck* is common in epistemology. If I am asked to guess what number you are thinking of and I happen to get it right, then (assuming I am not a mind reader) my guess is not an expression of knowledge, just dumb luck. Knowledge requires justification. However, even if I have a justified true belief, it might be undermined by the problem of epistemic luck. Moore (2012, 639) borrows an example from Bertrand Russell: I see my normally reliable clock read 5:40 and thereby come to hold the justified true belief that the time is 5:40. However, unbeknownst to me, the clock stopped working 24 hours ago, so it was pure luck that my belief turned out to be true. Here again, epistemic luck undermines knowledge.

Now, Moore applies the problem of epistemic luck to Robinson's account:

In the same way that Russell's broken clock is correct once a day, so the cause of a report of pain will be caused by a physical event that also causes pain in one possible world. But, in the same way that Russell's broken clock is also incorrect many times a day, and therefore is an unreliable time-telling device, so the cause of a report of pain will be caused by a physical event that also causes itchiness, tingliness, jealousy, et cetera, in many possible worlds, and so yields incorrect results in many possible worlds, and therefore is an unreliable epistemic mechanism. (Moore 2012, 639–640)

Moore's claim, then, is that *even if* we grant that $P_1 \rightarrow Q_1$ holds in the actual world, our report about Q_1 cannot express knowledge since it is a matter of epistemic luck that $P_1 \rightarrow Q_1$ holds; there are countless other possible worlds where it does not hold.

This claim strikes me as absurd for the following reason: If we take Moore's claim seriously, we cannot make knowledge claims about, say, the motion of the planets, since there are possible worlds in which the laws of gravity do not hold. Further, we could not consider *any* clock – however stable and precise – a reliable epistemic mechanism since there are possible worlds in which that particular clock is broken. Once again, Moore's objection runs too far in the direction of radical skepticism. It targets not just epiphenomenalism but any claim to knowledge that relies on physical or psychophysical laws.

If, as Moore grants, the psychophysical law $P_1 \rightarrow Q_1$ holds in this world, then it provides a reliable epistemic mechanism such that we can have knowledge of our qualitative events. The fact that there are nomologically distinct worlds where such a reliable mechanism does not exist is no threat to Robinson's view. Thus, Moore's third objection fails.

6. Conclusion

The self-stultification objection undoubtedly presents a challenge for epiphenomenalism, but not an insurmountable one. The objection assumes a causal requirement of knowledge of qualitative events, and I have argued that Robinson (2006) provides a plausible alternative to such a requirement by appealing to a common underlying cause. Further, I have defended this account against objections raised by Moore (2012). In doing so, I have only scratched the surface of a vast philosophical debate. But hopefully, I have shown that epiphenomenalism can hold its ground against the self-stultification objection and remain a plausible view about the relationship between mental and physical events.

References

- Block, Ned. 2002. "Concepts of Consciousness." *Philosophy of Mind: Classical and Contemporary Readings*, edited by David John Chalmers, 206-218. New York: Oxford University Press.
- Chalmers, David John. 2004. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

- Jackson, Frank. 2002. "Epiphenomenal Qualia." *Philosophy of Mind: Classical and Contemporary Readings*, edited by David John Chalmers, 273-280. New York: Oxford University Press.
- Moore, Dwayne. 2012. "On Robinson's Response to the Self-Stultifying Objection." *Review of Philosophy and Psychology* 3, no. 4: 627-41.
- Nida-Rümelin, Martine, and Donnchadh O'Conaill. 2019. "Qualia: The Knowledge Argument." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2019 Edition. <https://plato.stanford.edu/entries/qualia-knowledge/>.
- Robinson, William. 2006. "Knowing Epiphenomena." *Journal of Consciousness Studies* 13, nos. 1-2: 85-100.
- Robinson, William. 2019. "Epiphenomenalism." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2019 Edition. <https://plato.stanford.edu/entries/epiphenomenalism/>.
- Rosenthal, David M. 1997. "A Theory of Consciousness." *The Nature of Consciousness*, edited by Ned Block, Owen J. Flanagan, and Guven Guzeldere, 729-753. Cambridge, Mass: MIT Press.