



Rejecting epiphobia

Umut Baysan¹

Received: 23 March 2020 / Accepted: 10 October 2020 / Published online: 21 October 2020
© The Author(s) 2020

Abstract

Epiphenomenalism denies some or all putative cases of mental causation. The view is widely taken to be absurd: if a theory can be shown to entail epiphenomenalism, many see that as a *reductio* of that theory. Opponents take epiphenomenalism to be absurd because they regard the view as undermining the evident agency we have in action and precluding substantial self-knowledge. In this paper, I defend epiphenomenalism against these objections, and thus against the negative dialectical role that the view plays in philosophy of mind. I argue that nearly in all cases where a theory implies one kind of epiphenomenalism, it is an epiphenomenalism of a non-problematic kind, at least as far as issues about agency and self-knowledge are concerned. There is indeed a problematic version of epiphenomenalism, but that version is not relevant to the debates where its apparent absurdity is invoked.

Keywords Mental causation · Epiphenomenalism · Agency · Self-knowledge · Physicalism

1 Introduction

Epiphenomenalism denies some or all putative cases of mental causation. Of course, for epiphenomenalism to be an interesting view, its denial of mental causation must be *systematic* and *non-trivial*. It must be systematic, for if you allow that mental states/events/properties generally cause various sorts of events, but insist that your neighbour Frank's beliefs last weekend were causally inefficacious, that shouldn't make you an epiphenomenalist. By contrast, if you hold that Frank's and his conspecifics' beliefs and other intentional mental states can never cause anything, your denial of mental causation is systematic enough to warrant epiphenomenalism. The denial of mental causation must be non-trivial because if you are an eliminativist and reject the very existence of mental states/events/properties, you will also deny all puta-

✉ Umut Baysan
umut.baysan@philosophy.ox.ac.uk

¹ Merton College, University of Oxford, Merton Street, Oxford OX1 4JD, England, UK

tive cases of mental causation purely because you don't think there are any mental states/events/properties. But again, that should not make you an epiphenomenalist. As we shall shortly see, there are many ways of systematically and non-trivially denying mental causation, and hence being an epiphenomenalist.

In the contemporary literature, with a handful of exceptions, no one ardently defends epiphenomenalism.¹ Instead, the view often figures in *reductio* arguments in philosophy of mind. Consequently, many have sought to show that their views don't lead to epiphenomenalism, which has led to the development of subtly different accounts of mental causation and sophisticated theories of causation more generally. Fodor calls the general attitude in philosophy of mind towards this view “epiphobia”: the fear that one's view implies epiphenomenalism (1990: p. 137). Agreeing that epiphenomenalism must be avoided, Fodor proclaims that

if it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying ..., if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world (ibid: 156).

Similarly, Horgan thinks that epiphenomenalism is an “outrageous offense against common-sense” (1987: p. 504), suggesting that it “should be an utter last resort, to be embraced only if all viable alternatives prove to be even more paradoxical and untenable than epiphenomenalism itself” (ibid.). Others have called it “truly incredible” (McLaughlin 1994: p. 284), and suggested that “there is no real hope that epiphenomenalism could become a credible view” (Burge 1993: p. 112). But, what is so wrong about epiphenomenalism?

Although there are several arguments that have been advanced against epiphenomenalism, two of these stand out in suggesting not only that the view is false, but also that it is a good thing that it is false, as it has highly undesirable consequences. First, for those who think that agency in action requires mental causation, epiphenomenalism undermines agency in action, and hence presumably free will. In Kim's words, “we care about mental causation, ... chiefly because we care about human agency” (2009: p. 44). In his recent book-length defence of free will, List (2019) spends a whole chapter arguing against epiphenomenalism, as he thinks epiphenomenalism threatens freedom in action: “the relevant actions of any bearer of free will are caused, not merely by some nonintentional physical processes, but by the appropriate mental states” (ibid.: 24–25). Epiphenomenalism is then found problematic, as it is a non-negotiable fact that we are free agents of our actions. Second, epiphenomenalism is alleged to have unacceptable epistemological consequences. On the assumption that what is known must have some causal import, in particular on the knowing subject, epiphenomenalism implies that we can't know other minds, and more worryingly,

¹ See Robinson (2018) for a full-fledged defence of epiphenomenalism. In Robinson's use, “epiphenomenalism” refers to a positive claim about what mental phenomena *are* and how they *are* caused, as well as the negative claim that mental phenomena do *not* cause anything. Throughout the paper, I will take “epiphenomenalism” to refer to the negative claim which denies the causal efficacy of some or all mental phenomena. See Lyons (2006) and Tammelleo (2008) for other defences of epiphenomenalism in this negative sense.

we can't know even *our own minds* or mental states. As Moore puts it, “a necessary condition for the report that I was in pain to be counted as knowledge is that it was caused by the qualitative pain event itself” (2012: p. 630; see also De Brigard 2018). That we have such self-knowledge is very intuitive, hence epiphenomenalism seems to fly in the face of this intuitive fact.

I believe epiphobia draws its strength from these two arguments more so than others that we can find in the literature,² as these two arguments target our pre-theoretical intuitions about ourselves and our place in nature: we are free agents and we know our own minds. For this reason, in my assessment of epiphobia, I shall focus on these two arguments. I shall call these *the argument from agency* and *the argument from self-knowledge* respectively.

I am not an epiphenomenalist, but I am dissatisfied with the negative dialectical role epiphenomenalism has come to play in contemporary philosophy of mind. Thus, I propose to reassess the status of epiphenomenalism. To do this, I will explore answers to two intertwined questions. First, is epiphenomenalism really the undesirable view that the arguments from agency and self-knowledge seem to suggest? Second, do otherwise tenable views fall due to entailing epiphenomenalism? I will motivate largely negative answers to both of these questions.

The paper has two parts. In the first part (Sects. 2 and 3), I argue that there are various epiphenomenalist views, and most of these views are largely unaffected by the arguments from agency and self-knowledge. I argue for this by showing that the kinds of epiphenomenalism discussed in various debates in philosophy of mind are not really contrary to the intuitions that motivate these two arguments. In the second part (Sect. 4), I argue that some views that are often criticised for entailing a particularly worrisome kind of epiphenomenalism actually don't entail that kind of epiphenomenalism. They might imply the kinds of epiphenomenalism discussed in the first part of the paper, but as I will have shown, those kinds of epiphenomenalism don't fall prey to the arguments from agency and self-knowledge. The conclusion I draw from these is that epiphenomenalism doesn't deserve its negative dialectical role. This shows that some views that are adopted mainly because their alternatives lead to epiphenomenalism are not well-motivated. Throughout the paper, I will highlight examples of such views.

2 Varieties of epiphenomenalism

Let's start with understanding epiphenomenalism better. A very strong form of epiphenomenalism denies *any* kind of causal efficacy of *any* kind of mental state/event/property in relation to *any* kind of putative effect. Call this view “Epiphenomenalism” (with capital “E”).³ On this view, beliefs, desires, sensations, conscious experiences, and anything that populates one's mental life simply have no causal role in anything whatsoever.

² Other arguments against epiphenomenalism appeal to natural selection (Popper and Eccles 1977: pp. 73–4), metaphysics of causation (e.g. Menzies 2003; Kroedel 2019), properties (e.g. Baysan 2018), and metaphysics of mind–body relations (e.g. Shoemaker 2001). See Walter (2009a) for further discussion.

³ Henceforth, when “E” is capitalised, I mean to refer to this full-strength form of epiphenomenalism.

It goes without saying that Epiphenomenalism is a very strong view, and there are at least three ways one can weaken it and still retain an epiphenomenalist position. *First*, one might think that only some kinds of mental states/events/properties lack causal efficacy while others do have causal efficacy. For example, one might think that qualitative properties instantiated by conscious experiences are epiphenomenal, while intentional mental states are not. *Second*, one might believe that mental states/events/properties have causal efficacy in relation to further mental states/events/properties, but not in relation to the physical domain; perhaps conscious experiences cause other conscious experiences or intentional mental states, but they have no “downwards” causal powers. *Third*, one might take mental states/events/properties to have causal efficacy in one sense, while not in an allegedly more important sense. For example, one could argue that mental events are sometimes causes of some subsequent events, but they are not causes of these subsequent events *in virtue of* being mental events. To understand this idea, consider the sense of mental causation coined by Horgan as “mental quausation”: where mental states are causes *qua* mental states (Horgan 1989). So, on this third way of weakening Epiphenomenalism, we are considering views which accept mental causation, but deny mental quausation.

These three different ways of weakening Epiphenomenalism are orthogonal, hence there is an eight-way partitioning of the logical space for epiphenomenalist views:

- I. *Epiphenomenalism (with capital “E”)* No mental entity⁴ has any causal efficacy, in any sense in any domain.
- II. *Quausal epiphenomenalism* No mental entity has any causal efficacy in any substantial sense (though some mental entities may have causal efficacy in a non-substantial sense) in the physical or any other domain.⁵
- III. *Downward epiphenomenalism* No mental entity has any causal efficacy, in any sense, in the physical domain (though some mental entities may have causal efficacy in other domains).
- IV. *Downward quausal epiphenomenalism* No mental entity has any causal efficacy in any substantial sense (though some mental entities may have causal efficacy in a non-substantial sense) in the physical domain (though some mental entities may have causal efficacy in other domains).
- V. *Partial epiphenomenalism* Some (but not all) mental entities lack causal efficacy, in any sense, in any domain.
- VI. *Partial quausal epiphenomenalism* Some (but not all) mental entities lack causal efficacy in a substantial sense in any domain; though they may have causal efficacy in a non-substantial sense.
- VII. *Partial downward epiphenomenalism* Some (but not all) mental entities lack causal efficacy, in any sense, in the physical domain (though they may have causal efficacy in other domains).
- VIII. *Partial downward quausal epiphenomenalism* Some (but not all) mental entities lack causal efficacy in any substantial sense (though they may have causal efficacy in a non-substantial sense) in the physical domain; they may have causal efficacy (in the relevant sense) in other domains.

⁴ Henceforth, I use “entity” in a generic sense to cover states, processes, events, properties.

⁵ As with the term “mental quausation”, I borrow “quausal epiphenomenalism” from Horgan (1989: p. 51).

In the contemporary mental causation debate, Epiphenomenalism (I)—the strongest epiphenomenalist position on our list—is usually portrayed as an implausible consequence of *non-reductive physicalism* and (anti-physicalist) *property dualism*.⁶ Non-reductive physicalism is the variety of physicalism according to which mental properties are not identical with physical properties.⁷ Property dualism is similar to non-reductive physicalism in denying the reducibility of mental properties to physical properties, but differs from it as it rejects physicalism. It is often argued that these views imply Epiphenomenalism based on the consideration that only physical properties can be causally relevant or efficacious. The argument in support of this consideration is the topic of Sect. 4.

Some views in philosophy of mind are criticised for entailing quausal epiphenomenalism (II) or the related views of downward quausal epiphenomenalism (IV), partial quausal epiphenomenalism (VI), and partial downward quausal epiphenomenalism (VIII). To illustrate these, I will focus on one widely-discussed example of partial downward quausal epiphenomenalism. Recall that this is the type of view whereby (only) some kinds of mental states lack downward quausal efficacy. For this, consider anomalous monism, defended by Davidson (1970). On this view, propositional attitudes are identical with physical events, but this identity holds only at the *token* level. Propositional attitudes are not *type*-identical with physical events because different tokens of the very same propositional attitude type could fall under different physical event types. Davidson’s argument for this token-identity thesis proceeds by laying out the following claims: (i) some propositional attitudes cause physical events; (ii) causation requires subsumption under a strict causal law; and (iii) there are no strict causal laws that subsume propositional attitudes and physical events. These three claims lead to a puzzle which is solved thanks to the claim that every propositional attitude that is a cause of a physical event is identical with a physical event. Since there are strict causal laws that subsume physical events, propositional attitudes (which are identical with physical events) can be causes of physical events. Anomalous monism surely doesn’t imply Epiphenomenalism, as it is a part of the view that mental events cause physical events. But there is a worry that anomalous monism fails to explain an important sense in which mental events cause physical events. The worry is that, on this view, mental events cause physical events, but *not* in virtue of being mental events; instead, they cause physical events in virtue of being physical events. That is, mental events are not causes *qua* mental, but they are causes *qua* physical; there may be mental causation but there is no mental quausation.⁸ Hence, if such worries are well-founded, we have an example of partial downward quausal epiphenomenalism. Those who think this kind

⁶ Since both views are traditionally characterised as views about mental *properties*, I will take the kind of epiphenomenalism that they are alleged to imply to be epiphenomenalism about mental *properties*.

⁷ Thus, the so-called “token-identity” theory (discussed in the next paragraph) is *not* a form of reductive physicalism according to my construal.

⁸ See Sosa (1984), Horgan (1989), and McLaughlin (1989). The same point is sometimes articulated in a slightly different ontological language. For example, sometimes, the worry is that, in Davidson’s view, there is mental causation, but mental *properties* do not have any role in this; all the causal work is done by physical properties. I find this way of presenting this point misleading, as Davidson doesn’t present anomalous monism in terms of properties. In Davidson’s view, “it is events that have causes and effects ... [I]t makes no literal sense to speak of an event causing something ... by virtue of its mental properties” (1993: p. 13).

of epiphenomenalism is unacceptable think that token-identity is not enough, and we should hold a type-identity theory (see Kim 1989, 1998: p. 37).

Consider downward epiphenomenalism (III) and partial downward epiphenomenalism (VII) next. Articulated in terms of properties, downward epiphenomenalism is the view that mental properties don't have "downwards" causal powers: they may have causal powers in relation to events that involve property instantiations at the mental level, but not at the physical level. For example, beliefs, desires, and conscious experiences may cause other mental phenomena, but they don't cause any physical phenomena. This view assumes that there are different levels of reality, that the mental level is a higher-level compared to the physical level, and also that mental properties are not identical with physical properties. A notable example of a view which posits downwards causal powers for mental properties is emergentism (see McLaughlin 1992; to be discussed in Sect. 4 below).

A much-discussed example of partial downward epiphenomenalism is the view that is associated with Jackson (1982): *qualia*, namely the properties of some mental states in virtue of which there is something it's like to have them, don't have downwards causal powers.⁹ On this view, qualia are non-physical properties. With the assumption that the physical domain is causally closed (a point we will visit in Sect. 4), this view is said to imply that qualia can't have physical effects. Why is this claim restricted to downward causation? The reason is that the defenders of this view accept that qualia may have effects insofar as these effects are not in the physical domain (ibid.: p. 133). For example, the reddish quale of my visual experience could causally contribute to my awareness of a red object in front of me (where such awareness is a mental state), but it doesn't have any causal role with respect to the physical properties of my nervous system. Those who think that it is absurd that qualia are epiphenomenal in this sense look for ways of defending a physicalist account of qualia (see Horgan 1987).

Qualia epiphenomenalism needn't be understood as an example of downward epiphenomenalism. One could simply hold that qualia don't have any causal powers whatsoever, in which case qualia epiphenomenalism would be an example of partial epiphenomenalism (V). Another interesting example of partial epiphenomenalism, at least as far as some opponents of the view are concerned, is the kind of picture that is implied by externalism about mental content (henceforth, externalism). Roughly, externalism is the view that contents of some intentional mental states don't supervene on the intrinsic properties of the bearers of such intentional states. Some mental states are such that possessing them requires being appropriately related to certain things in one's environment (Putnam 1975; Burge 1979). For example, if the transparent, odourless, liquid substance you call "water" is not actually water but a different substance, XYZ, the thoughts that you express by using the word "water" are not about water. The kind of content that doesn't supervene on intrinsic properties is *wide content*. Those who think that causation is an intrinsic matter and the causal powers of an entity supervene on its intrinsic properties (Fodor 1991) argue that externalism implies a kind of epiphenomenalism: mental states with wide content don't have causal powers (partial epiphenomenalism); or if they do, not in virtue of their content (partial causal epiphenomenalism). Those who think that this sort of epiphenomenalism is unaccept-

⁹ I discuss qualia epiphenomenalism in detail in Baysan (2019b).

able either try to show ways of explaining the causal relevance of wide content (see Dretske 1990) or explore other notions of mental content (see Fodor 1991).¹⁰

Having presented various examples of epiphenomenalist positions, I have so far passed no judgement on whether there is anything wrong with any of these views. In the next section, I will revisit the two central arguments against epiphenomenalism (the argument from agency and the argument from self-knowledge), and explore to what extent they are relevant to the epiphenomenalist positions we have seen in this section.

3 The arguments from agency and self-knowledge

3.1 The argument from agency

The argument from agency suggests that we should reject epiphenomenalism on the grounds that it deprives us of agency in action. The worry is that without being agents with the capacity to act, we wouldn't have free will, and without free will, we wouldn't have moral responsibility, thus epiphenomenalism seems to undermine free will and moral responsibility. Of course, this way of putting things assumes that free will requires agency, and that moral responsibility requires free will—I will take these assumptions for granted. My question in this section is: Does epiphenomenalism really deprive us of agency in action? The answer is not that simple, because we don't have one epiphenomenalism, but many.

Assuming that free will requires the capacity to act, and that this capacity requires some of our mental states to be causes of our actions, it seems to follow that the argument from agency is spot on with respect to Epiphenomenalism (with capital “E”). As previously noted, two views in the literature that allegedly imply Epiphenomenalism are non-reductive physicalism and non-physicalist property dualism, views that I will turn to in Sect. 4.

What about partial epiphenomenalism? This is the view—or better, the family of views—according to which only some kinds of mental states or properties are epiphenomenal. Whether the argument from agency is successful against partial epiphenomenalism depends on what part of mentality we are considering. As previously noted, two notable versions of partial epiphenomenalism are qualia epiphenomenalism and wide content epiphenomenalism. Let's consider both positions.

If qualia epiphenomenalism is true, then qualia are epiphenomenal properties. Qualia are properties of experiences in virtue of which such experiences are experiences—in virtue of which there is something it's like to have them. So, qualia epiphenomenalism is explicitly committed to the claim that “the properties in virtue of which we are conscious are not properties in virtue of which events causally contribute

¹⁰ Another partial epiphenomenalist position may hold that mental content in general—not only wide content—is causally irrelevant. Since wide content epiphenomenalism is motivated by the assumption that causation is an intrinsic matter—i.e. causal powers supervene on intrinsic properties—I don't know what could motivate such a strong version of partial epiphenomenalism, hence I will not discuss this position below. Suffice it to say that a view along these lines will be virtually indistinguishable from Epiphenomenalism, hence my remarks on Epiphenomenalism may easily apply to it.

to our behaviour” (Robinson 2020: p. 27). On this view, the *painfulness* of Hilda’s pain experience upon touching a very hot stove doesn’t have any causal role in her saying “ouch!”, or moving her hand away from the hot stove. Does this problematise or undercut Hilda’s capacity to act in anyway? Clearly not, as this is a case of a reflexive behaviour, and not a case of action.

But suppose, just a moment later, upon reflecting on her persisting painful experience, Hilda decides to run cold water on her hand, and then successfully executes this plan. If painfulness is epiphenomenal—as the qualia epiphenomenalist believes—then it doesn’t play any causal role in the successful execution of Hilda’s plan. From this, does it follow that Hilda’s running cold water on her hand is not a free act, or that Hilda is not the agent of this action? I think not, as it is not clear to me that the causal powers of qualia have much to do with having or lacking the capacity to act. It would be odd if Hilda’s pain experience had nothing to do with the fact that she runs cold water on her hand given that she seems to decide to do so upon having a pain experience. In general, it would be very undesirable if we were systematically wrong about the aetiology of our actions. But it is important to note that qualia epiphenomenalism doesn’t say that our *experiences* don’t have any causal role in our actions. It says that the *qualia* that are instantiated by our experiences don’t have any causal role. So, even if qualia epiphenomenalism is true, it can still be true that Hilda’s pain experience plays a causal role in her action. On the assumption that qualia epiphenomenalism is true, what is false is the claim that *the painfulness of this pain experience* is part of the causal story. It may turn out that a view whereby pain experiences are causes but painfulness doesn’t have causal role in this is false. But falsehood is not my question here; the question is whether this view undermines agency. I see no reason to believe that it does.

Am I assuming that an experience can be a cause without the qualia that are constitutive of that experience playing any causal roles? Well, I am assuming that, *according to qualia epiphenomenalism*, this can be—and indeed *is*—the case. But in so doing, I am also assuming that this is a coherent metaphysical picture. I want to stand by this assumption, because I think it is plausible that not all properties of causing events are causally relevant to their effects. Suppose I throw a brick at a window, and the window breaks. As it happens, the brick is red. Surely, the mass and the rigidity of the brick are causally relevant to the breaking; but not all properties of the brick are causally relevant to the breaking. So, some properties of the brick-throwing event (e.g., it being a red-brick-throwing event) are causally irrelevant, although the event itself is surely causally efficacious to the breaking in virtue of its causally relevant properties. The qualia epiphenomenalist—for reasons of her own—thinks that qualia are further examples of causally irrelevant properties of causally efficacious events.

Here, one might object that I am ignoring an important sense in which qualia of our experiences seem relevant to our actions. When Hilda provides reasons for her actions, she sometimes cites the qualitative aspects of her experiences. She says that the painfulness of her experience is why she is running cold water on her hand. Likewise, when she acts in ways such that she is pursuing pleasant experiences, her actions are motivated by the pleasantness of some past experiences that are similar to the ones that she is pursuing. If qualia are epiphenomenal, how can we make sense of these facts? Well, the answer is that *if qualia epiphenomenalism is true, these are not facts*.

What is important is that these not being facts doesn't undermine the fact that Hilda's mental states cause her action. If these are not facts, then there is a sense in which our claims to be *ideal* agents are seriously compromised. But the argument from agency doesn't rest on the assumption that we are ideal agents. And even if it did, so much the worse for it: we are often wrong about the reasons for our actions, we sometimes fail from the weakness of the will, and these already seem to show that we are not ideal agents. If qualia epiphenomenalism is true, then there is yet one more reason why we are not ideal agents. However, none of this means that we lack the capacity to act and free will.

The more relevant examples of mental states, as far as agency is concerned, are propositional attitudes. Consider the following simplified example. Upon hearing knocking, I believe that someone is at the door, and want to find out who is at the door; the pairing of my belief and desire causes me to open the door. Should it turn out that beliefs and desires can never be causes, scenarios like this may never obtain. We have seen that one example of partial epiphenomenalism holds that intentional mental states with wide content are epiphenomenal. On the assumption that externalism is true, beliefs and desires are standard examples of intentional mental states that have wide content. Then, one might think that this kind of partial epiphenomenalism is undermined by the argument from agency.

But the problem is merely apparent. If wide content epiphenomenalism were to be understood as the view that the beliefs and desires that are the *vehicles* of wide content are epiphenomenal, on the view that all beliefs and desires have only wide content, wide content epiphenomenalism would undermine our capacity to act as agents. But there is no reason to understand wide content epiphenomenalism as a view about the vehicles of wide content. In fact, if we were to understand vehicles of content in the standard sense, namely as “physical particulars that bear contents and *whose causal interactions explain behaviour*” (Shea 2018: p. 15, emphasis added), wide content epiphenomenalism, understood as a view about vehicles of wide content, would be self-contradictory.

Suppose that externalism is true for intentional mental states, and my belief that there is water in the bathtub is not a belief that my Twin Earth doppelgänger (around whom there is XYZ, not water) shares with me. Instead, my doppelgänger believes that there is XYZ in the bathtub. Suppose, upon believing that there is water in the bathtub, I put my toddler in the bath. If it is true that this view implies that wide content is epiphenomenal, it follows from this view that the fact that the content of my belief corresponds to water (and not to XYZ) has no causal role whatsoever in my putting my toddler in the bath. Something that does *not* follow from this view is that my belief, i.e. the vehicle of the relevant content, has no causal role in my putting the poor creature in the bath. Again, this view may be false, but it is not obvious at all that it implies that my beliefs are not causes of my actions.

As I suggested in Sect. 2, there is reason to formulate wide content externalism as a quausal epiphenomenalist view: my belief that there is water in the bathtub causes me to put my toddler in the bath, but not in virtue of being a belief about water. Let's consider the quausal versions of epiphenomenalism in relation to the argument from agency.

I believe none of the quausal variants of epiphenomenalism fall prey to the argument from agency; at least, not very easily. There is no reason to treat the four versions of quausal epiphenomenalism separately, as I think the same point applies to all. For us to lack the capacity to act, for us to cease to be agents, what is required is that our mental states fail to cause our actions. If my beliefs and desires are causes of my actions, but not in virtue of being beliefs and desires (or not in virtue of having the relevant wide content), I still have the capacity to act, my mental states still cause my actions, and hence I am the agent of my actions. In other words, mental causation is sufficient for action, mental quausation is optional.

In the forgoing discussion—throughout the last five paragraphs—I have assumed that it is the pairings of beliefs and desires that cause actions, and such causings can be sufficient for agency in action. One might worry that this assumption is too weak—i.e. too easy to satisfy. In particular, if one holds that the relevant mental states that cause action must also *provide reasons* for actions (see Davidson 1963), then it is not clear that causation by the vehicles of mental content alone can satisfy the requirement for agency. After all, how can a pure vehicle—without reference to a content—provide reason for an action?¹¹ Likewise, if one holds, with Mele (1992), that “our desires, beliefs, and the like ... help explain behaviour ... at least partly *in virtue of their content*” (ibid.: p. 11, emphasis added), and moreover that such explanation is a requirement for agency in action, then it is probably false that mental causation *without* mental quausation can ever be sufficient for agency in action. From this, it follows that this particular epiphenomenalist position—i.e. wide content epiphenomenalism understood as a quausal epiphenomenalist view—is in tension with the so-called “standard” causal theory of action which subscribes to these claims.¹² This is bad news for wide content epiphenomenalism insofar as the argument from agency is presented with the assumption of the causal theory of action in the background. Upon reflection, this shouldn’t come as a surprise: the causal theory of action seems to require exactly what wide content epiphenomenalism and quausal epiphenomenalism more generally explicitly reject: mental quausation by content. Therefore, although the argument from agency can be mounted against wide content epiphenomenalism, wide content epiphenomenalists can resist it by rejecting the standard causal theory of action.

Hence, putting Epiphenomenalism aside, I have argued that epiphenomenalist views or epiphenomenalist implications of some views are largely unaffected by the argument from agency. I shall now turn to the argument from self-knowledge.

3.2 The argument from self-knowledge

The main epistemological worry about epiphenomenalism that I shall focus on is about the possibility of self-knowledge, or knowledge of one’s own mind. An important background assumption in presenting this problem is that knowledge requires causal relevance, so what is known has some causal relation to the knowing agent’s knowledge (Goldman 1967: pp. 358–9). The worry is this: if our mental states don’t have any

¹¹ Thanks to an anonymous referee for pressing me on this point and putting this problem in these words.

¹² See also Aguilar and Buckareff (2010) for a helpful discussion of the causal theory of action.

causal relevance, how do we get to know them? The argument from self-knowledge is motivated by the intuitive power of this worry.

Assuming either knowledge is a mental state or a knowing subject's beliefs that partially constitute her knowledge are mental states, if self-knowledge requires mental causation, it requires same-level causation, not downward causation. Because of this, I will set downward version of epiphenomenalism aside. I will also leave Epiphenomenalism aside (until Sect. 4). That leaves us three versions to consider: partial epiphenomenalism (in particular qualia epiphenomenalism and wide content epiphenomenalism), quausal epiphenomenalism, and partial quausal epiphenomenalism.

Let's start with qualia epiphenomenalism as an example of partial epiphenomenalism. The worry is that qualia epiphenomenalism implies that the knowledge of qualia is not possible: I can't know the qualia of my own experiences. Read one way, this is very dramatic. If we equate the "qualia" talk with "what it's like" talk, it might even sound like this view suggests that I can't know what it's like for me to see something red when I see something red. However, as I shall argue, even coupled with the background assumption of a causal theory of knowledge, qualia epiphenomenalism doesn't imply any of this.

There are two different ways of understanding qualia and how they might relate to our knowledge of them. On the first understanding, qualia are simply properties of our experiences in virtue of which there is something it's like to have these experiences. This understanding is neutral on whether we are meant to be acquainted in our consciousness with the very properties (or instances thereof) that are responsible for our consciousness. So, on this understanding, knowledge of our own experiences doesn't require any form of direct acquaintance with the qualia that are constitutive of our experiences. In other words, this understanding of qualia leaves it open as to whether qualia are directly or immediately apprehensible in consciousness. Therefore, even if knowledge of an experience requires a causal relation that one bears to that experience, such knowledge does not *require* the causal efficacy of qualia that are constitutive of that experience.

On a second understanding, qualia are not only the properties of our experiences in virtue of which there is something it's like to have these experiences, but they are also the properties that we are directly acquainted with in our experiences. Dennett (1988), in his critique of the very concept of "qualia", argues that qualia must be properties of this kind. He says: "*since* they are properties of my experiences ... qualia are essentially directly accessible to the consciousness of their experiencer ... or qualia are properties of one's experience with which one is intimately or directly acquainted" (ibid.: p. 385, emphasis added). If qualia, understood this way, don't have causal powers, how could we be acquainted with them? How could they cause our acquaintance of them? If they can't cause our acquaintance of them, then we can't be acquainted with them. This is absurd, and in fact paradoxical, given that we are defining qualia partially as properties that we are directly acquainted with.

I concede that *if* qualia are to be understood this latter way, and direct acquaintance requires causal connection, *then* we would have an untenable combination of views. Qualia epiphenomenalism would not be only implausible, but also paradoxical. However, for this very reason, it would be odd for a defender of qualia epiphenomenalism

to understand qualia as direct objects of our acquaintance *and* to also hold a causal theory to explain how we are acquainted with our qualia.

In response to this formulation of the argument from self-knowledge, qualia epiphenomenalists can do either of the following two things. First, they can reject this second way of understanding qualia and simply hold that qualia are properties of our experiences in virtue of which we have conscious experiences, but they are not properties that we are acquainted with. I believe this is a coherent view, because I don't think Dennett's inference—quoted above—is a valid one. From the fact that F is a property of my experience E, it doesn't follow that by having E, I am directly acquainted with F. To be clear, on this view, examples of qualia are not going to be sensory properties, such as phenomenal redness. But this is an acceptable result, as phenomenal redness is not a property *of* an experience—in the same way that redness is not a property *of* a picture of a red tomato. (Some things that partially constitute the picture of a red tomato are red, but the picture itself is not red.) On this view, when a subject experiences phenomenal redness R, there is something it's like for her to experience R, and this is partially because her experience instantiates a quale Q (where $Q \neq R$).

Second, qualia epiphenomenalists—should they take qualia as properties that we are directly acquainted with—can reasonably help themselves to a non-causal account knowledge of qualia, or phenomenal knowledge. They can, for example, accept Nagasawa's (2010) view that such acquaintance is a *constitutive* relation, rather than a causal one: a subject's "phenomenal knowledge about qualia *q* is partly constituted by *q*" (ibid: 52). On this view, we might want to require a causal connection when explaining our knowledge of objects in our external environment, but we don't need to do this in order to explain our knowledge of our own experiences. Whatever the merits of this account of phenomenal knowledge might be, it is clear that it is an alternative to the causal account, and it helps qualia epiphenomenalists maintain a non-paradoxical position.

What about wide content epiphenomenalism? If the kind of epiphenomenalism that is supposedly implied by externalism is true, then, with the assumption of the causal theory of knowledge, it follows that we don't always know the contents of some of our intentional mental states. But consider an alternative option whereby mental states have both wide content and *narrow* content, where the latter is the kind of content that supervenes on the intrinsic properties of the bearers of the relevant mental state. Following this alternative option, it is possible to argue that when one has a particular belief, one knows only the narrow content of the belief. I am not in the business of defending this alternative option; I simply want to note that wide content epiphenomenalists who want to resist the argument from self-knowledge have a coherent way of doing so.

Finally, let's consider the two relevant versions of quausal epiphenomenalism together. Applied to quausal versions of epiphenomenalism, the epistemological worry is that *if* mental states or properties can't be causally efficacious in virtue of being mental, *then* we couldn't know, or be acquainted with, them. If this conditional claim is true, then quausal epiphenomenalism and its variants imply scepticism about our own minds. The obvious route for quausal epiphenomenalists is through denying this conditional claim. Even if a causal theory of knowledge is presupposed, from this presupposition alone it doesn't follow that what is known must be known in virtue of

all of its properties; some properties of the objects of our knowledge may be causally inefficacious. If quausal epiphenomenalism is true, then examples of such causally efficacious properties include properties such as *being mental*: some (or all) mental states are such that their *being mental* is not causally efficacious to how they are known.

All in all, full-strength Epiphenomenalism aside, there is no reasonable version of epiphenomenalism that is refuted by the argument from self-knowledge. This, I say, with the assumption that the knowledge of one's own mental states and properties requires a causal relation. Of course, if one were to drop this assumption, the argument from self-knowledge is unfounded even against full-strength Epiphenomenalism.

4 Non-reductive physicalism, property dualism, and Epiphenomenalism

Having argued that relatively weak epiphenomenalist positions cannot be easily dismissed based on the arguments from agency and self-knowledge, I will now turn to an assessment of Epiphenomenalism, the view which denies mental causation altogether. As I have highlighted several times, if agency and self-knowledge are worries for any kind of epiphenomenalism, they should be worries for Epiphenomenalism.

In the contemporary mental causation debate, Epiphenomenalism notably features in the causal argument for *reductive physicalism*: namely the view mental properties are identical with physical properties. In a nutshell, the argument is that only physical properties can be causally efficacious in a physical world, so if mental properties are not physical properties, then they can't be causally efficacious. Kim's (1998, 2005) articulation is the paradigmatic version of this argument, so I will focus on this version. This is the so-called "exclusion argument" for its purported conclusion that any putative non-physical mental cause is "excluded" by some purely physical cause.

The exclusion argument is leveraged against two alternatives to reductive physicalism: non-reductive physicalism and (non-physicalist) property dualism. The argument has the following structure: on the assumption that there is mental causation (i.e. Epiphenomenalism is false), commitments of non-reductive physicalism and property dualism, together with further assumptions, lead to inconsistent positions. To avoid such inconsistency, non-reductive physicalists and property dualists can endorse Epiphenomenalism, but Epiphenomenalism is an unacceptable position. Hence, these views ought to be rejected, and the alternative view, reductive physicalism ought to be accepted. The relevant mutually inconsistent claims are the following five:

- (1) *Mental causation* Some mental properties, or instantiations thereof, are causally sufficient for some events. (In other words, Epiphenomenalism is false.)
- (2) *Causal closure* Every event that has a cause at a time has a sufficient physical cause at that time. This is true not only for events at the physical level, but also events that supervene on physical events.¹³

¹³ Sometimes, a weaker version of this claim is endorsed. This weaker claim says that every *physical* event that has a cause at a time has a sufficient physical cause at that time. In the arguments in which this weaker version is used, the target cases of mental causation are merely cases of downward causation. Since, in this section, I am discussing all putative cases of mental causation, not merely downward causation, I will work with this stronger claim of causal closure.

- (3) *Exclusion* No event has multiple simultaneous sufficient causes unless it is causally overdetermined—i.e. if there are two or more individually sufficient simultaneous causes of an event, that event is causally overdetermined.
- (4) *No overdetermination* Cases of mental causation are not cases of causal overdetermination. For if they were, then the world would contain systematic causal overdetermination. There may be rare occasions of causal overdetermination, but there is no systematic causal overdetermination.
- (5) *Distinctness* Mental properties are not reducible to, or identical with, physical properties.

Let me illustrate the inconsistency of these five claims through our previous example. Upon touching the hot stove, Hilda’s pain experience causes her to move her hand. In line with (1) above, we have a mental cause (pain experience) and an effect (hand movement). As (2) dictates, at the time at which Hilda’s behaviour has a cause, it has a sufficient physical cause. Let’s call this sufficient cause “SPC”. Given (5), Hilda’s pain experience (or the relevant mental property in virtue of which she has this experience) is not identical with SPC (or the relevant physical property in virtue of which SPC obtains). So far, these three suggest that there is an event (hand movement), which has multiple simultaneous sufficient causes (a mental property and a physical property, or instances thereof). If (3) is true, then this must be a case of causal overdetermination. But if (4) is true, this can’t be a case of causal overdetermination. Therefore, claims (1)–(5) are mutually inconsistent. Non-reductive physicalists and property dualists can achieve consistent positions if they reject (1); hence, these views lead to Epiphenomenalism.

The problem with this argument is that it is not clear why non-reductive physicalists and property dualists must accept *all* of the remaining claims (2)–(5). Since they are not reductive physicalists, they have to accept (5). But the remaining options are possible candidates, and in what follows, I shall argue that non-reductive physicalists and property dualists *could* (and perhaps *should*) reject (3) and (2) respectively.

Property dualists, insofar as they are non-physicalists, needn’t accept (2). Whatever their arguments in favour of a non-physicalist position might be, once their view admits physicalism-violating, irreducible, *sui generis*, mental properties, it is not clear why they should embrace a principle that is assumed in defending physicalism. A relevant example of a property dualist position is emergentism (e.g. Broad 1925; O’Connor 1994). According to a standard conception of emergentism (see McLaughlin 1992; Wilson 2015), this view is committed to a particularly strong downward causation principle: some higher-level properties “influence motion in ways unanticipated by laws governing less complex kinds and conditions concerning the arrangements of particles” (McLaughlin 1992: p. 51). This principle not only posits downward causation whereby some higher-level goings-on can cause some physical goings-on, but also implies a violation of the causal closure of the physical domain. If some higher-level properties causally influence physical goings-on in ways that physical properties may not (as implied by the principle), then it would seem that the causal efficacy of some higher-level properties goes beyond the causal efficacy of physical properties. The upshot of this is that a venerable property dualist position straightforwardly opposes the idea portrayed in (2), in which case it doesn’t follow that property dualists must be

cornered into Epiphenomenalism in order to achieve an internally coherent position. In other words, property dualism doesn't have to lead to Epiphenomenalism.¹⁴

Unlike property dualists, non-reductive physicalists are not in a position to reject (2). But they can reject (3). Non-reductive physicalists, in virtue of being physicalists, are committed to a metaphysically necessary connection between mental properties and the physical properties they supervene on. For if mental properties don't metaphysically supervene on physical properties, the case for physicalism is undercut (see Chalmers 1996). A case against (3) is that once two putative simultaneous causes of a subsequent event are necessarily connected such that one necessitates the other, they are not distinct enough to generate causal overdetermination.

An argument by Bennett (2003) illustrates this last point as follows. For two events, A and B, to causally overdetermine E, it has to be *non-vacuously* true that (i) if A hadn't happened but B had happened, then E would still have happened, and (ii) if B hadn't happened but A had happened, then E would still have happened. This, after all, is how we understand causal overdetermination in standard "firing-squad" cases: Shooter 1 and Shooter 2 simultaneously shoot Victim; each shot is lethal; if Shooter 1 hadn't fired, Shooter 2 would have, and Victim would have still died; if Shooter 2 hadn't fired, Shooter 1 would have, and Victim would have still died. Now, if A is metaphysically necessitated by B, it is not metaphysically possible for B to occur without A occurring, which implies that if condition (i) is true, it can only be *vacuously* true. The resulting picture generates a counterexample to (3) because the non-reductive physicalist can say that Hilda's pain experience and SPC are individually sufficient simultaneous causes of her behaviour, but they don't causally overdetermine it; it is metaphysically impossible for SPC to obtain without Hilda's pain experience obtaining.¹⁵

Now, this response to (3)—and more generally to the exclusion argument—is controversial. As Morris (2015) argues, this so-called "disanalogy-style" response to the argument rests on a problematic dialectic. The counterfactual analysis of overdetermination given in conditions (i) and (ii) above is based on standard "text-book" cases of overdetermination (such as the firing-squad cases), yet the response ultimately boils down to the claim that the mental-physical cases are disanalogous to such text-book cases. Moreover, making this point doesn't seem to require to say much about the metaphysics of mind–body relations, and this is a further worry raised by critics of the disanalogy-style response. As Morris argues, "it seems that the [exclusion] problem has been dismissed without any discussion of either mental causation or the metaphysics of non-reductive physicalism" (ibid.: p. 443).¹⁶ Given these worries, non-reductive

¹⁴ It needs to be acknowledged that property dualists *can*—and indeed sometimes *do*—accept (2). In fact, Robinson's (2019) Epiphenomenalist brand of property dualism partially rests on the causal closure of the physical. In Robinson's case, it would be dialectically odd to argue against his view for implying Epiphenomenalism, as his view is *ardently* Epiphenomenalist. I am merely suggesting that rejecting (2) is a possible option for property dualists who are compelled by the "exclusionist" reasoning.

¹⁵ Though, see Keaton and Polger (2014) for problems raised against this latter claim.

¹⁶ Morris (2015) also argues that there is a problem with formulating the exclusion problem with (3) in the first place, which seems to have counterexamples that have nothing to do with the mental-physical cases. Counterexamples of this kind can be found in Árnadóttir and Crane (2013). He is right that the "exclusionist" reasoning could be presented by different arguments which wouldn't inherit the objectionable features of (3). As said above, I am focusing on Kim's paradigmatic formulation of this problem which endorses (3) explicitly (e.g., Kim 2005: p. 42).

physicalists would do better to look for a way of responding to the exclusion argument by illuminating the nature of the relationship between the mental and the physical.

Fortunately, there is a viable non-reductive physicalist position whereby (3) comes out false because of the way mental properties are related to the physical properties that realize them (and not merely because there are independent counterexamples to it). An account of the relationship between mental properties and their realizers that I find particularly promising is given by what has come to be known as the “subset view” (Wilson 1999, 2011 Shoemaker 2001, 2007).¹⁷ According to the subset view, a relationship that holds between a mental property, *M*, and its physical realizer, *P*, is that the causal powers of *M* are a proper subset of the causal powers of *P*. If this is true, the relevant causal powers of Hilda’s pain experience are ultimately “inherited” from the causal powers of SPC. (Here, I am assuming that SPC is a realizer of Hilda’s pain experience.) According to Shoemaker’s (2001) version of this theory, this implies that Hilda’s pain experience is a proper part of SPC. Since it is implausible that wholes and their parts causally overdetermine their effects, we have a counterexample to (3).

In Wilson’s version of this theory, we don’t have a part-whole relation, but we still have an explanation of why Hilda’s pain experience and SPC don’t causally overdetermine their effects. According to Wilson, an implication of the proper subset relationship between *M*’s and *P*’s causal powers is the following: when *M* is instantiated in virtue of *P*’s instantiation, every causal power of *M* is token-identical with some causal power of *P*. When Hilda’s pain experience causes her hand movement and SPC causes her hand movement, there is only one relevant causal power that is exercised on this very occasion. Given that only one causal power is being exercised, it makes little sense to think that *M* and *P* causally overdetermine *E*. Therefore, (3) is false.

So far so good. But doesn’t this type of view still imply Epiphenomenalism? If *M*’s causal powers are inherited from *P*, wouldn’t it be more appropriate to say that, really, *P* is the cause, and *M* is epiphenomenal? This position might imply some kind of epiphenomenalism, but it doesn’t imply Epiphenomenalism. The worry here seems to be about quausation, not causation: if *M*’s causal powers are ultimately causal powers of physical properties, how could *M* have any causal efficacy in virtue of being mental? So, even if these worries may lead to some kind of epiphenomenalism, the kind of epiphenomenalism we get is at best *quausal* epiphenomenalism. My question here is whether we get Epiphenomenalism (with capital “E”) from non-reductive physicalism; I have argued that we don’t. *If* we get quausal epiphenomenalism however, that is also fine, as I have argued in Sect. 3 that quausal epiphenomenalism is not as implausible as is commonly assumed, as it doesn’t easily fall prey to the arguments from agency and self-knowledge. So, either way, non-reductive physicalism is on solid ground—as far as mental causation is concerned anyway.

To conclude this section, non-reductive physicalism and property dualism don’t imply Epiphenomenalism. Non-reductive physicalists don’t have to accept the claim that mental properties and their realizers causally overdetermine their effects; and

¹⁷ For objections to the subset view, see Gillett (2003), McLaughlin (2007), Morris (2011, 2013), Noordhof (2013), and Pineda and Vincente (2017). For more recent defences of the core claim of the subset view, see Baysan (2016, 2018, 2019a—which includes replies to some of the aforementioned critiques). See also Walter (2009b) for more general worries about “realization-based” strategies for saving non-reductive physicalism from Epiphenomenalism.

property dualists don't have to accept the causal closure of the physical. If they want to accept these claims and the further claims that fuel the exclusion argument, then Epiphenomenalism becomes a *choice*. And if it is a choice, then it makes little sense to argue against these views on the basis of an epiphobic *reductio*.

5 Conclusion

Epiphenomenalism is often found problematic—*absurd* even—as it is thought to undermine the evident agency we have in action and preclude substantial self-knowledge. Focusing on the arguments from agency and self-knowledge, I have distinguished between various kinds of epiphenomenalism, and argued that most of these views cannot be easily dismissed based on these two arguments. This of course doesn't mean that these epiphenomenalist positions are true or ultimately defensible, as there may be other reasons why they may not be tenable. But it is important to note that they don't stand or fall with agency and self-knowledge. This is a significant result, as I believe the alleged *absurdity* of epiphenomenalism is motivated by pre-theoretical intuitions that support the arguments from agency and self-knowledge. I have also argued that there is indeed an epiphenomenalist view that stands out as problematic: the full-strength Epiphenomenalist position that denies mental causation altogether. But I showed that two notable views that are criticised for implying Epiphenomenalism in fact don't imply it. Based on these results, my conclusion is that epiphenomenalism doesn't deserve the negative dialectical role it has been given.

Acknowledgments Many thanks to Dominic Alford-Duguid, Katherine Baysan, Jonas Christensen, Simone Gozzano, Fiona Macpherson, Neil McDonnell, Michele Paolini Paoletti, Ema Sullivan-Bissett, Nathan Wildman, Jessica Wilson, David Yates, members of his research project Emergence in Natural Sciences (funded by Fundação para a Ciência e a Tecnologia: PTDC/FER-HFC/30665/2017), and three (excellent!) referees for *Synthese* for their comments on previous versions of this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aguilar, J. H. & Buckareff, A. A. (2010). The causal theory of action: origins and issues. In J. H. Aguilar & A. A. Buckareff (eds.), *Causing human actions: New perspectives on the causal theory of action* (Oxford: Oxford University Press, pp. 1–26).
- Árnadóttir, S. & Crane, T. (2013). There is no exclusion problem. In S. C. Gibb, E. J. Lowe, & Ingthorsson (eds.), *Mental causation and ontology* (Oxford: Oxford University Press, pp. 248–266).
- Baysan, U. (2016). An argument for power inheritance. *The Philosophical Quarterly*, 66(263), 383–390.
- Baysan, U. (2018). Epiphenomenal properties. *Australasian Journal of Philosophy*, 96(3), 419–431.
- Baysan, U. (2019a). Emergence, function, and realization. In S. C. Gibb, R. F. Hendry, & T. Lancaster (eds.), *Routledge handbook of emergence* (Abingdon: Oxford, pp. 77–86).

- Baysan, U. (2019b). Mad Qualia. *Philosophical Quarterly*, 69(276), 467–485.
- Bennett, K. (2003). Why the exclusion problem seems intractable, and how, just maybe, to tract it. *Noûs*, 37, 471–497.
- Broad, C. D. (1925). *The mind and its place in nature*. London: Routledge & Kegan Paul.
- Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy*, 4, 73–122.
- Burge, T. (1993). Mind-body causation and explanation. In J. Heil and A. Mele (eds.), *Mental causation* (Oxford: Clarendon, pp. 97–120).
- Chalmers, D. (1996). *The conscious mind. In Search of a fundamental theory*. Oxford: Oxford University Press.
- Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, 60, 685–700.
- Davidson, D. (1970). Mental events. In L. Foster & J. W. Swanson (Eds.), *Experience and theory* (Amherst (pp. 79–101). MA: University of Massachusetts Press.
- Davidson, D. (1993). Thinking causes. In J. Heil and A. Mele (eds.), *Mental causation* (Oxford: Clarendon, pp. 3–18).
- De Brigard, F. (2018). Defence of the self-stultification objection. *Journal of Consciousness Studies*, 21(5–6), 120–130.
- Dennett, D. (1988). Quining qualia. In A. J. Marcel & E. Bisiach (Eds.), *Consciousness in contemporary science*. Oxford: Oxford University Press.
- Dretske, F. (1990). Précis of explaining behavior: Reasons in a world of causes. *Philosophy and Phenomenological Research*, 50(4), 783–786.
- Fodor, J. (1990). *A theory of content and other essays*. Cambridge: MIT Press.
- Fodor, J. (1991). A modal argument for narrow content. *Journal of Philosophy*, 88, 5–26.
- Gillett, C. (2003). The metaphysics of realization, multiple realizability, and the special sciences. *Journal of Philosophy*, 100, 591–603.
- Goldman, A. (1967). A causal theory of knowing. *Journal of Philosophy*, 64, 355–372.
- Horgan, T. (1987). Supervenient qualia. *The Philosophical Review*, 96, 491–520.
- Horgan, T. (1989). Mental quausation. *Philosophical Perspectives*, 3, 47–76.
- Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly*, 32, 127–136.
- Keaton, D., & Polger, T. (2014). Exclusion, still not tracted. *Philosophical Studies*, 171, 135–148.
- Kim, J. (1989). The myth of nonreductive materialism. *Proceedings and Addresses of the American Philosophical Association*, 63(3), 31–47.
- Kim, J. (1998). *Mind in a physical world*. Cambridge: MIT Press.
- Kim, J. (2005). *Physicalism or something near enough*. Princeton: Princeton University Press.
- Kim, J. (2009). Mental causation. In B. McLaughlin, A. Beckermann, and S. Walter (eds.), *The Oxford handbook of philosophy of mind* (Oxford: Clarendon, pp. 29–52).
- Kroedel, T. (2019). *Mental causation: A counterfactual theory*. Cambridge: Cambridge University Press.
- List, C. (2019). *Why free will is real*. Cambridge: Harvard University Press.
- Lyons, J. (2006). Defense of epiphenomenalism. *Philosophical Psychology*, 19, 767–794.
- McLaughlin, B. (1989). Type epiphenomenalism, type dualism, and the causal priority of the physical. *Philosophical Perspectives*, 3, 109–135.
- McLaughlin, B. (1992). The rise and fall of british emergentism. In A. Beckerman, H. Flohr, & J. Kim (Eds.), *Emergence or reduction? Prospects for nonreductive physicalism* (pp. 49–93). Berlin: De Gruyter.
- McLaughlin, B. (1994). Epiphenomenalism. In S. Guttenplan (ed.), *A Companion to the philosophy of mind* (Oxford: Blackwell, pp. 277–288).
- McLaughlin, B. (2007). Mental causation and shoemaker-realization. *Erkenntnis*, 67, 149–172.
- Mele, A. R. (1992). *Springs of action: Understanding intentional behavior*. New York: Oxford University Press.
- Menzies, P. (2003). The causal efficacy of mental states. In Walter and Heckmann (eds.), *Physicalism and mental causation: The metaphysics of mind and action* (Exeter: Imprint Academic, pp. 195–223).
- Moore, D. (2012). On Robinson's response to the self-stultifying objection. *Review of Philosophy and Psychology*, 3, 627–641.
- Morris, K. (2011). Subset realization, parthood, and causal overdetermination. *Pacific Philosophical Quarterly*, 92, 363–379.
- Morris, K. (2013). On two arguments for subset inheritance. *Philosophical Studies*, 163, 197–211.
- Morris, K. (2015). Against disanalogy-style responses to the exclusion problem. *Philosophia*, 43, 435–453.
- Nagasawa, Y. (2010). The knowledge argument and epiphenomenalism. *Erkenntnis*, 72, 37–56.

- Noordhof, P. (2013). Mental causation: Ontology and patterns of variation. In S. Gibb, E.J. Lowe & R.D. Inghorsson (eds.), *Mental causation and ontology* (Oxford: Oxford University Press, pp. 88–125).
- O'Connor, T. (1994). Emergent properties. *American Philosophical Quarterly*, 31, 91–104.
- Pineda, D., & Vincente, A. (2017). Shoemaker's analysis of realization: A review. *Philosophy and Phenomenological Research*, 94, 97–120.
- Popper, K., & Eccles, J. (1977). *The self and its brain*. New York: Springer.
- Putnam, H. (1975). The meaning of "Meaning". *Minnesota Studies in the Philosophy of Science*, 7, 131–193.
- Robinson, W. (2019). *Epiphenomenal mind: An integrated outlook on sensations, beliefs, and pleasure*. New York: Routledge.
- Robinson, W. (2020). Thick NCCs yield physicalist epiphenomenalism. *Journal of Consciousness Studies*, 27(7–8), 77–94.
- Shea, N. (2018). *Representation in cognitive science*. Oxford: Oxford University Press.
- Shoemaker, S. (2001). Realization and mental causation. In C. Gillett & B. Loewer (Eds.), *Physicalism and its discontents* (pp. 74–98). Cambridge: Cambridge University Press.
- Shoemaker, S. (2007). *Physical realization*. Oxford: Oxford University Press.
- Sosa, E. (1984). Mind-body interaction and supervenient causation. *Midwest Studies in Philosophy*, 9, 271–281.
- Tammelleo, S. (2008). A limited defense of epiphenomenalism. *South African Journal of Philosophy*, 27, 40–51.
- Walter, S. (2009a). Epiphenomenalism. In B. McLaughlin, A. Beckermann, & S. Walter (eds.), *The oxford handbook of philosophy of mind* (Oxford: Oxford University Press, pp. 85–94).
- Walter, S. (2009b). Taking realization seriously: No cure for epiphobia. *Philosophical Studies*, 151, 207–226.
- Wilson, J. (1999). How superduper does a physicalist supervenience need to be? *The Philosophical Quarterly*, 49, 33–52.
- Wilson, J. (2011). Non-reductive realization and the powers-based subset strategy. *The Monist*, 94, 121–154.
- Wilson, J. (2015). Metaphysical emergence: Weak and strong. In T. Bigaj & C. Wuthrich (Eds.), *Metaphysics in contemporary physics: Poznan studies in the philosophy of the sciences and the humanities* (Leiden: Brill, pp. 251–306).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.