

Causal Closure, Synaptic Transmission and Emergent Mental Properties



Giacomo Zanotti

Abstract The causal argument for physicalism about the mind has received a lot of attention. In particular, the literature has focused on the main premise of the argument, namely the causal closure principle (CC). In this article, I present and discuss the so-called argument from physiology, that is widely regarded as the most convincing line of reasoning in favour of CC. When it comes to providing empirical grounds for the argument from physiology, the most promising move the physicalist can opt for is to focus on the mechanisms of synaptic transmission. Here, I argue that the argument from physiology can provide support for CC only if evidence concerning synaptic transmission is combined with a non-innocent assumption about the internal causal organisation of the nervous system. I contend that this assumption should be vindicated. Unfortunately, this does not seem to be possible at the moment.

Keywords Causal closure principle · Physicalism · Emergent mental properties · Argument from physiology · Mental causation

1 The Causal Argument for Physicalism

Physicalism about the mind is arguably the prevailing position in the contemporary debate on the nature of mental states. Yet, it is not a unitary one. First, a distinction must be made between identity theories and physicalist views that take mental properties to be nothing over and above physical properties without positing (type) identities. Most notably, however, things get complicated when it comes to providing a more precise characterisation of nothing-over-and-aboveness. For this purpose, advocates of physicalism have resorted to different metaphysical relations, ranging from realisation (Mehnyk 2006) to constitution (Pereboom 2011) and grounding (Dasgupta 2014) – this list is not meant to be exhaustive – but the question is still

G. Zanotti (✉)
Politecnico di Milano, Milano, Italy
e-mail: giacomo.zanotti@polimi.it

open. True, some common denominators can be found. In particular, starting with Lewis (1983), it has been argued that all physicalists share at least the commitment to the claim that mental properties are metaphysically supervenient on physical properties.¹ That said, a consensus on a precise definition of physicalism is still lacking.

Dualism is the traditional alternative to physicalism. In this article, I will focus on *property* dualism, and more specifically on its emergence-based formulations. According to the advocates of emergentist dualism, mental properties are *strongly* emergent from physical properties: they are fundamental, only nomologically – as opposed to metaphysically – necessitated by physical properties.² In addition, strongly emergent mental properties are often taken to have fundamentally novel causal powers, not possessed by their physical emergence bases (McLaughlin 1992; O'Connor 1994; Kim 1999).

In what follows, I will often refer to the contraposition between physicalism and dualism in terms of weak versus strong emergence.³ Weakly emergent mental properties are acceptable from a physicalist perspective: they are non-fundamental properties, metaphysically necessitated by physical properties, and lacking novel causal powers. Notice that I am admittedly oversimplifying. A detailed discussion of physicalism and dualism would require a long detour I cannot afford. However, the characterisations I have provided should allow us to address the main subject of this article, that is the causal argument for physicalism.

Among the arguments for physicalism, the causal one has probably received the most attention in the literature.⁴ The line of reasoning hinges on the incompatibility between a dualist view and the following premises:

¹This has been disputed (Montero 2013; Montero and Brown 2018; see Alter 2021 for a reply). However, the greatest majority of authors take metaphysical supervenience to be the physicalist's minimal commitment.

²Wilson (2015, 2021) calls into question this way of conceiving strong emergence, arguing that there could be cases in which properties that we would tend to regard as strongly emergent are metaphysically necessitated by physical properties – e.g. a Malebranchean scenario in which God always causes the instantiation of fundamental mental properties upon the occasion of physical properties. However, for our purpose, a modality-based definition of strong emergence will do the work (among others, see Chalmers 2006; Noordhof 2010).

³To be clear, these are not the only possible solutions to the mind-body problem. In particular, panpsychism has been receiving a lot of attention in the recent literature (see Bruntrup and Jaskolla 2016; Goff 2017). However, the debate seems to be still largely driven by the dichotomy between physicalism and dualism. This is also the picture emerging from the latest PhilPapers survey (<https://survey2020.philpeople.org>), in which 51.9% of the participants leaned towards physicalism, 32.1% towards anti-physicalism, and 15.9% opted for 'other'. A more specific question on consciousness in the survey reveals that 22% of the participants leaned towards dualism, while only 7% were sympathetic to panpsychism.

⁴I focus on the line of reasoning that is discussed in the contemporary debate (see Papineau 2001, 2002). However, it is worth highlighting that the causal argument has some antecedents in the history of Western philosophy. It is sufficient to think about Princess Elisabeth of Bohemia writing to Descartes that it would be easier 'to concede matter and extension to the soul than to concede the capacity to move a body and to be moved by it to an immaterial thing' (III 685 AT, in Shapiro 2007,

- (1) mental properties have physical effects;
- (2) all physical effects are fully caused by purely *physical* prior histories (the causal closure principle);⁵
- (3) the physical effects of mental properties are not systematically overdetermined by physical properties.⁶

If all three of these claims are true, then dualists are in serious trouble. On the one hand, they hold that *sui generis* mental properties, metaphysically distinct from the physical properties they emerge from, are among the causes of our behaviours. On the other hand, (2) implies that our behaviours already have a history of *sufficient* physical causes. The only way to ease this tension would be to argue that our behaviour is systematically overdetermined by mental and physical causes. However, this possibility is precluded by (3).

To facilitate the discussion, let us consider a simplified case of mental causation in which a mental property M and a physical property P compete in the production of a physical event *e*.⁷ Four options seem to be available to the dualist when presented with the causal argument:

- A. First, the dualist could reject (1). This way, M would turn out to be merely epiphenomenal. No overdetermination would be involved since M would have no causal power.
- B. Otherwise, the dualist could reject the causal closure principle. Clearly, if there is no constraint on the nature of the events that can be part of the sufficient causal history of *e*, then there is nothing wrong with M playing an *ineliminable* role in the production of *e*.⁸ Again, no overdetermination would be involved, since no physical property P would compete with M.
- C. A third possibility is to deny (3). Both M and P would count as individually sufficient causes of *e*, that would turn out to be genuinely overdetermined. Note that in this case, unlike in (B), *e* would occur even if M – or P – failed to be

p. 68). Interestingly, the same line of reasoning can be found in Lucretius' *De rerum natura*, III 162–168.

⁵This version of the principle is employed in Papineau (2002). I take a closer look at the possible formulations of the principle in the next section.

⁶According to the standard definition of overdetermination, A and B overdetermine an event *e* iff:

- (i) A and B are distinct events;
- (ii) A is sufficient for causing *e*;
- (iii) B is sufficient for causing *e*;
- (iv) If A did not occur, *e* would still occur;
- (v) If B did not occur, *e* would still occur.

⁷In the debate on causal closure, the standard notion of event is Kim's (1976) one, according to which events are 'property exemplifications' that can be represented as ordered triples $\langle x, P, t \rangle$. Focusing on monadic events, an event consists of an object *x* instantiating a property *P* at a moment *t*.

⁸By *ineliminable*, I mean that if M failed to be instantiated, then *e* would not occur. No other property would be instantiated in place of M, making up for the absence of M's causal powers.

instantiated. The instantiation of the remaining property would be sufficient for the instantiation of *e*.

- D. Lastly, the dualist could reject property dualism and admit that the relevant causal powers of M simply coincide with the relevant causal powers of P – the easiest way to do it is probably to admit that M and P are at least token-identical. In this case, M could still be taken to be the cause of *e* without violating (2), given that mental causation would turn out to be just physical causation.

Although there are four possible ways out from the inconsistency, the physicalist takes for granted the truth of premises (2) and (3) of the causal argument, significantly reducing the range of moves the dualist can opt for. As a result, the conclusion of the argument comes in the form of a dilemma: either mental properties are merely epiphenomenal, or dualism is false. Needless to say, neither option is desirable for the dualist.

In discussing the causal argument, I will not consider the possibility of claiming that mental and physical properties overdetermine their effects in all cases in which a given physical event is supposed to be caused by a conscious occurrence. True, it has been argued that systematic overdetermination in cases of mental causation is not particularly problematic.⁹ However, in case M and P overdetermined their effect *e*, *e* would occur even if M failed to be instantiated. On the contrary, the dualist I have in mind holds that mental properties play an *ineliminable* role in the production of our behaviour. Arguing that *sui generis* mental properties have causal powers that systematically overdetermine their effects seems to be a metaphysically onerous, ad hoc move. For these reasons, I will just assume (3) along with the physicalist. My aim is rather to show that the rejection of (2) is a viable option for the dualist.

I start by outlining the two competing models of mental causation the physicalist and the dualist appeal to (Sect. 2). After that, I focus on the causal closure principle. After arguing for a specific formulation of the principle (Sect. 3), I consider the arguments that have been provided in its favour (Sect. 4). In particular, I focus on Papineau's argument 'from physiology'.¹⁰ I argue that the evidence from neuro-physiology Papineau has in mind does not provide direct reasons in favour of the causal closure principle. A further assumption concerning the internal causal organisation of our nervous system is needed. However, such an assumption may be harder to justify than the causal closure principle itself (Sect. 5).

Before proceeding, however, let me say something about the rejection of (1). The dualist's first reaction to the causal argument could be to bite the bullet and accept

⁹See Sider (2003); a different view is defended in Bernstein (2016). The literature on overdetermination is vast and complex. In addition, what is usually at stake in the debate is the possibility of non-reductive, physicalist models of overdetermining mental causation. As far as I can see, arguing for systematic overdetermination within a dualist framework is way more difficult.

¹⁰Note that this is not the only argument we have available. In particular, the argument 'from fundamental forces' (Papineau 2001) is frequently discussed in the literature. However, the argument from physiology is 'broadly considered much more convincing' (Dimitrijević 2020), and it seems to be the one even Papineau insists upon in his latest contributions (see Papineau 2020).

that *sui generis* mental properties have no causal powers upon the physical domain. This view, which falls under the label of epiphenomenalism, has an illustrious history. Among others, Malebranche's occasionalism and Leibniz's pre-established harmony can be regarded as theoretical prototypes of epiphenomenalism. Both reject the possibility of any causal interaction between the mind – or, more precisely, the soul – and the body. More recently, epiphenomenalism has been defended by philosophers such as Campbell (1970), Jackson (1982), and Robinson (2019).¹¹ Chalmers (1996) expressed a certain sympathy for non-interactionist forms of dualism as well, although he is more cautious in his later works (see Chalmers 2010).

Interestingly, it is sometimes suggested that there is some empirical evidence in favour of epiphenomenalism. More specifically, reference is made to Libet's (1985) famous experiments. To make a long story short, these experiments would show that some neural activations, that would be responsible for the initiation of simple movements, significantly precede the conscious decision to perform those movements. Therefore, physical properties would pre-empt mental properties of their causal role. As far as I can see, there are good reasons for being suspicious. First, it should be kept in mind that some methodological aspects of these studies have been disputed (Gomes 2002, Pockett and Purdy 2011; see also Lavazza 2016). Furthermore, the philosophical implications of the obtained results are far from being clear (Mele 2014; Baumeister et al. 2018).

In addition to this, some arguments against epiphenomenalism have been provided in the literature. Among others, the one from natural selection and the self-stultification objection are worth recalling. In broad brushstrokes, the former is a line of reasoning leveraging the intuition that, if consciousness were epiphenomenal, then its evolution would be inexplicable (Popper and Eccles 1977).¹² The latter, instead, is an argument to the effect that epiphenomenalism would be simply incompatible with our knowledge of our own mental states (see De Brigard 2014 for a discussion). True, advocates of epiphenomenalism could easily challenge the evolutionary argument. In particular, they could argue that the evolution of consciousness is 'a sort of byproduct' of physical evolution (Chalmers 2010, p. 131). Resisting the self-stultification objection, however, may be more complex.

For space reasons, I cannot afford to go into details. Let me just add that, even if compelling counter-arguments to the self-stultification objection were provided, giving up the causal efficacy of consciousness would not be so easy. Epiphenomenalism is at odds with our basic intuitions about the way our behaviour is influenced by our psychological life. Here is how Fodor (1989, p. 77) puts it:

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.

¹¹ See also Baysan (2020).

¹² See also James (1879) on the evolutionary utility of pleasure and pain.

Note that I am not assuming that our intuitions are always infallible in the context of theory choice in metaphysics, nor that they are in this specific case. All I am pointing out is that the implications of epiphenomenalism are extremely counterintuitive. In what follows, I will assume that the dualist has good reasons for avoiding the epiphenomenalistic apocalypse. In fact, I take the causal efficacy of mental states to be a *desideratum*, regardless of the metaphysical theory one ends up adopting.

With these premises, we can proceed with our analysis. In the next section, I will briefly present and discuss the two different models of mental causation the physicalist and the dualist are committed to.

2 Two Models of Mental Causation

As we have seen, four claims are at stake in the causal argument. I have assumed that both the physicalist and the dualist agree on two points:

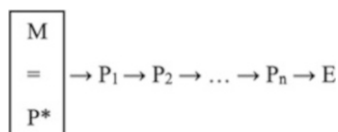
- causal efficacy: mental properties have physical effects;
- no overdetermination: the physical effects of mental properties are not overdetermined by physical properties.

What they disagree upon are the following claims:

- causal closure: all physical effects are fully caused by purely physical prior histories;
- property dualism: mental properties are something over and above physical properties.

To make things easier, let us consider a simple scenario in which a subject *S* is thirsty, and their experience of thirst (*M*) seems to be the cause of a chain of physical events (P_1, P_2, \dots, P_n) – motor neurons firing, sarcomeres contracting, and so on – that ultimately result in *S* ingesting water (*E*). Include another physical state P^* , that co-varies with *M* and can be referred to as *M*'s emergence base. What is at stake in the argument is the precise nature of the relation between *M* and P^* .

The physicalist avoids the inconsistency among the premises of the causal argument by rejecting (property dualism). Here is a schematic illustration of the physicalist model of mental causation – the arrows represent causal relations, going from causes to effects:

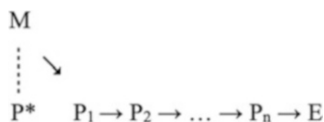


According to the physicalist, M and P* are at least token-identical, and the distinction that is made is purely conceptual.¹³ As a result, there are no two sources of causal power.

On the contrary, the dualist – or at least, the dualist I have in mind – aims at maintaining both (property dualism) and (causal efficacy). In their view, M and P* are still instantiated together. However, they are metaphysically distinct properties. M strongly emerges from P*, and the co-variation between them is accounted for in terms of contingent psychophysical laws. In addition, there should be room for a causal arrow that goes from M to P₁. Clearly, the plausibility of this model depends on the possibility of rejecting the causal closure principle.

At this stage, the physicalist could react by arguing that this project is utterly unrealistic, even conceding that (causal closure) may not hold. It could be argued that if both M and P* alone were sufficient for P₁ and were instantiated at the same time, (no overdetermination) would be violated. What is more, even if overdetermination worries are mitigated – e.g. by adopting a non-oomphy notion of causation *à la* Lewis/Woodward – independent considerations make this model implausible. If we admit that M and P* co-occur at *t* and that P* alone is enough for bringing about P₁, why should we posit another sufficient cause? There is a sense in which assigning a causal role to M appears to be a metaphysically onerous, dispensable move.

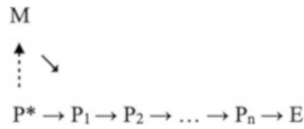
The dualist can easily address the issue by pointing out that, with (causal closure) out of the picture, nothing forces us to presuppose that the exercise of P*'s causal power is sufficient for causing P₁, or even that P* has a causal role at all. Once (causal closure) is rejected, dualists have two options. First, they can argue that M is the only cause of P₁:



Arguably, this sounds like heresy to the physicalist. However, once (causal closure) is dismissed, there is nothing absurd in the hypothesis of a mixed causal chain, that presumably starts before M with a series of physical – and possibly mental – events, involves M, and finally ends with E.

¹³Whether non-reductive versions of physicalism are ruled out by the causal argument is a controversial issue that has been extensively debated. In particular, Kim's (1998, 2005) exclusion argument against non-reductive physicalism is worth mentioning. I will not go into the details of Kim's line of reasoning, for which a number of non-reductive solutions have been suggested (among others, see Bennett 2003). In what follows, I will concede that the physicalist can reject dualism on the grounds of the causal argument without being committed to a strong identity thesis between mental and physical properties.

Alternatively, the dualist can argue that M and P^* are co-causes of P_1 . That is, the causal powers of both M and P^* are required to bring about P_1 . In particular, I have in mind the model defended in Lowe (2000, 2003):



In Lowe's view, P^* is the physical base for the emergence of M . More precisely, the instantiation of P^* is the *cause* of the instantiation of M at t . Crucially, both M and P^* count as causes of P_1 at t . However, since they are taken to be co-causes, they do not overdetermine P_1 . Neither P^* nor M can bring about P_1 alone, and P_1 would not occur in case M or P failed to be instantiated at t .

In the rest of this article, I will focus on the contraposition between physicalism and emergentist dualism *à la* Lowe, that seems to me to be the most promising interactionist model of mental causation the dualist can resort to. In both cases, P^* , that is the physical property serving as emergence base for M , plays a causal role with respect to P_1 . What is at issue, besides the precise nature of the emergence relation between M and P^* , is whether P^* 's causal powers are sufficient for the occurrence of P_1 or the contribution of a *sui generis* mental property is also needed.

3 The Causal Closure Principle

At this stage, we are in the position to carefully consider the causal closure principle, starting with its formulation. As pointed out in the literature, one of the major difficulties the physicalist must deal with is to provide an *adequately strong* version of the principle (Lowe 2000).¹⁴ If the adopted formulation is too weak, the causal argument fails to provide reasons for believing that physicalism is true – or, more precisely, that interactionist dualism is false. If the principle is too strong, on the other hand, it turns out to be almost indistinguishable from the conclusion of the causal argument. In addition, intuitively, the stronger is the principle, the harder it is to vindicate.

When providing an outline of the causal argument, I referred to the causal closure principle as the claim that 'all physical effects are fully caused by purely *physical* prior histories'. This formulation, however, is to a large extent ambiguous. In particular, it is not entirely clear what 'fully' means in this context. On the one hand, one could interpret the principle as stating that, at every moment t in the causal history of a given physical effect e , if e has a cause at t , then e has a sufficient physical cause at t . On the other hand, one could take the principle to state that the

¹⁴For an overview of the provided formulations, see Gibb (2015).

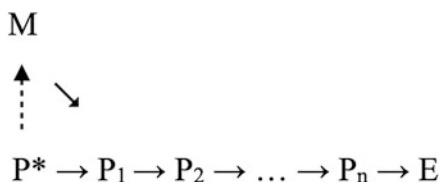
causal histories of physical events are *exclusively* made up of physical events. Let us take a closer look at these two different readings.

The first interpretation is arguably the most common in the debate. I will refer to it as the *weak* formulation of the causal closure principle (WCC):

(WCC) if a physical event e has a cause at t , it has a sufficient physical cause at t .¹⁵

The first thing to be noticed is that WCC is stronger than another similar principle that can be found in the literature, according to which ‘if a physical event has a cause that occurs at t , it has a physical cause that occurs at t ’ (Kim 2005, p. 43). If the sufficiency requirement is not specified, then the principle is way too permissive with respect to the acceptability of causally efficacious *sui generis* mental properties. Most notably, it is compatible with the possibility that both physical and non-physical events are needed at a moment t to bring about a given physical effect. Consider the dualist scenario in which M – a *sui generis* mental property – is instantiated at t together with its emergence base P^* . Suppose that both M and P^* are required for the production of a given physical effect e . In this model, even if M plays an ineliminable causal role in the production of e , e has a physical cause at t , so the causal closure principle without the sufficiency requirement is respected. Clearly, this is not enough for the physicalist, who aims at ruling out *sui generis* mental causation.

WCC, on the contrary, is explicit about the fact that e has a *sufficient* physical cause at t . Prima facie, if M and P^* were co-causes of e in the way just described, P^* would not count as a sufficient cause of e at t , and WCC would be violated. Apparently, if WCC holds, the only way for M to somehow cause e would be to be an *overdetermining* cause of it. The problem is that, on closer inspection, WCC too turns out to be compatible with a dualist, interactionist view of mental causation. Let us consider Lowe’s emergentist model of mental causation I have outlined in the last section. As we have seen, in Lowe’s view, the instantiation of a physical property P^* at t simultaneously causes the emergence of a metaphysically distinct mental property M . Importantly, M and P^* are co-causes of P_1 at t :



¹⁵Specifying t is needed to rule out the possibility that a physical event e occurring at t_2 has a purely mental cause M at t_1 that is in its turn the effect of a physical cause P at t . If the causal closure principle simply stated that ‘all physical effects have sufficient physical causes’ (as in Papineau 1998), e could have a purely mental cause at t_1 and still respect the principle. Since causation is (usually regarded as) a transitive relation, the fact that P causes M at t would be enough for granting that e has a sufficient physical cause in the scenario just described.

In this case, even if M plays an ineliminable role in the production of P_1 , P^* transitively counts as a sufficient cause of P_1 at t . P^* is sufficient for causing M at t , that together with P^* is sufficient for causing P_1 . Unless we reject the possibility of simultaneous causation, there seems to be room for an interactionist dualism that respects WCC.

This is something the physicalist is not willing to accept. While it is hardly questionable that Lowe's model is compatible with WCC, it is clear that it fails to respect the *spirit* of the causal closure principle. What the physicalist really has in mind when resorting to the causal closure principle is arguably something along these lines:

Pick any physical event [...] and trace its causal ancestry or posterity as far as you would like; the principle of causal closure of the physical domain says that this will never take you outside the physical domain. (Kim 1998, p. 40)

This directly leads us to the second interpretation of Papineau's (2002) formulation, that is the *strong* causal closure principle (SCC):

(SCC) physical events can only have physical causes

Clearly, this prevents the dualist from opting for interactionist models of mental causation *à la* Lowe. Since M is not a physical property, there is no room for the exercise of its causal powers in the causal history of P_1 , full stop. Admittedly, SCC may sound appealing. However, the physicalist should be cautious in adopting it in the context of the causal argument for physicalism. If the formulation of the principle is too strong, the causal argument begs the question. In fact, it is not difficult to see how this can happen.

One of the most insidious issues the physicalist must deal with when providing a formulation of the causal closure principle concerns the meaning of 'physical' (Crane and Mellor 1990). In order not to fall prey to Hempel's dilemma, physicalists often opt for a negative definition, according to which 'physical' should be interpreted as 'non-mental'.¹⁶ Hence, SCC turns out to be equivalent to the following principle:

(SCC*) physical events can only have non-mental causes

SCC*, in its turn, can be rewritten as the claim that *sui generis* mental properties cannot cause physical events. This, however, is suspiciously close to the conclusion the physicalist aims at reaching by means of the causal argument. If not question-begging, the causal argument would turn out to be redundant. In particular, once the possibility of all kinds of mental-to-physical causation is ruled out, it is not clear what role the no-overdetermination premise should have.

I take WCC and SCC to be examples of how different formulations of the causal closure principle can be too weak or too strong. In what follows, I will adopt the following one:

¹⁶In the sense of not fundamentally mental. Most notably, this view has been defended by Montero and Papineau (2005).

(CC) if a physical event e has a sufficient cause, it has an immediate sufficient physical cause¹⁷ (Papineau 2009)

On the one hand, CC is sufficiently strong. The immediacy requirement is meant to rule out the possibility of models *à la* Lowe, in which M is a sort of causal intermediary between P^* and P_1 , and P^* (partly) causes P_1 indirectly by simultaneously causing M .¹⁸ For CC to hold, P^* must be a sufficient cause of P_1 and *directly* cause P_1 , without the intermediate intervention of M . On the other hand, CC is not *too* strong. The possibility of *sui generis* mental properties having a causal role is still open. Most notably, they could be overdetermining causes of the physical effects CC refers to. To reach the desired conclusion, the physicalist must combine CC with the rejection of systematic overdetermination. Therefore, including CC among the premises of the causal argument does not make the line of reasoning a question-begging one.

At this stage, a precise formulation of the causal closure principle has been provided. In the remainder of this article, I will focus on the way the physicalist argues for its truth.

4 The Argument from Physiology

The arguments in favour of CC that are usually discussed in the literature are the argument ‘from fundamental forces’ and the argument ‘from physiology’ (Papineau 2001, 2002). In a nutshell, the former line of reasoning is an inductive one that insists on the fact that a number of *prima facie* special forces turned out to be reducible to a limited set of fundamental, conservative physical forces.¹⁹ The conclusion is that there are no special mental forces that are irreducible to basic physical forces. The argument from physiology, instead, hinges on the fact that despite the impressive progress in recent physiological research, no trace of special mental forces has been found. On the contrary, physical explanations for a number of biological – and more specifically, neural – phenomena have been provided. The conclusion, once again, is that there is no room for special mental forces. Admittedly, both arguments are not conclusive. However, this is not necessarily a problem for the physicalist. After all, there are many beliefs we entertain without having conclusive reasons for doing

¹⁷Note that this does not make mediated causation *per se* problematic. Given a physical event e_1 and its alleged physical effect e_2 , CC is perfectly compatible with the possibility of an intermediate physical event e^* that is caused by e_1 and causes e_2 . What CC rules out is the possibility of *non-physical* causal intermediacy.

¹⁸See also Garcia (2014) on ontologically proximal and distal causes.

¹⁹The argument from fundamental forces should not be confused with the one suggested by Dennett (1991), according to which the exercise of special mental forces would violate the principle of conservation of energy. Despite its initial appeal, this line of reasoning seems to be irremediably flawed; see Papineau (2002), Gibb (2010), Tomasetta (2015).

it. Rejecting the arguments in question solely because of the fact that they are not knock-down ones does not seem fair.

In these pages, I will focus on the argument from physiology. This choice is due to the fact that this line of reasoning is generally considered more convincing. Among other things, unlike the argument from fundamental forces, it does not require one to make the question-begging assumption that mental forces – whatever they might be – are not fundamental (Garcia 2014).

To get started, let us look at the formulation of the argument from physiology provided by Papineau (2001, p. 27):

[...] there is no direct evidence for vital or mental forces. Physiological research reveals no phenomena in living bodies that manifest such forces. All organic processes in living bodies seem to be fully accounted for by normal physical forces.

As far as I can see, two distinct – although inevitably interrelated – components of the argument can be isolated. On the one hand, the emphasis is on the fact that we have failed to detect any kind of action in living bodies that could be ascribed to special mental forces operating. On the other hand, scientists have succeeded in providing physical explanations for a number of biological phenomena and processes. Note that, in both cases, the evidence at stake is arguably the one that is provided by research in neurobiology and neuroscience broadly conceived. After all, if there is something that is affected by *sui generis* mental forces, that is arguably the nervous system. In what follows, I will consider the two components of the argument individually.

4.1 *The First Component*

Let us start with the claim that physiological research has provided no evidence for the action of *sui generis* mental forces. Referring to the latest advancements in biological sciences, Papineau (2001, p. 31) argues:

[...] these developments made it difficult to go on maintaining that special forces operate inside living bodies. If there were such forces, they could be expected to display some manifestation of their presence. But detailed physiological investigation failed to uncover evidence of anything except familiar physical forces.

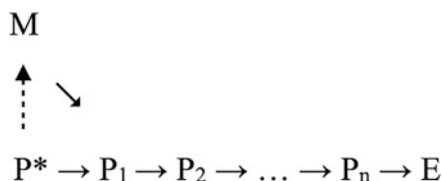
A way to resist this line of reasoning immediately comes to mind. The claim that there are no special mental forces, the dualist may argue, cannot be legitimately inferred from the lack of evidence attesting to such forces' action. The absence of evidence would be regarded as evidence of absence, and this is a typically fallacious move.

I am suspicious about dismissing the line of reasoning in question by simply branding it as a case of *argumentum ad ignorantiam*. If one thinks about it, there seem to be cases in which the absence of evidence can provide reasons for believing that something is actually non-existent. Montero (2003) explicitly addresses this point by discussing the example of ghosts. There is a sense in which the absence of

evidence for ghosts' existence can provide reasons for believing that they actually do not exist. The condition is that we also have knowledge of what really causes noises echoing in the night and the other phenomena that could be traced back to ghosts' action and could have led us to posit their existence in the first place.

Similarly, Montero argues, the absence of evidence for the existence of *sui generis* mental forces can serve as evidence of absence, provided that 'we also have a fairly good understanding of what fundamentally nonmetal force actually causes us to cry out when in pain, and so forth' (2003, p. 185).²⁰ This sounds pretty reasonable. Hence, it all seems to come down to whether we know enough of the physical processes that are supposed to be the causes of our behaviour. Interestingly, this issue is of the utmost importance also when it comes to assessing the second component of the argument from physiology. For this reason, I will leave the question unanswered for the moment. What I would like to do, now, is to focus on a couple of potential problems that pertain exclusively to the first component.

A crucial assumption of the first component of the argument from physiology is that special mental forces would be empirically detectable. If this were not the case, then appealing to the absence of evidence for their efficacy would be preposterous. The fact that scientists have never observed *sui generis* mental causation is perfectly compatible with the existence of empirically undetectable mental forces. Unfortunately for the physicalist, however, it seems that we cannot exclude the possibility that mental properties' causal contribution is undetectable. Let us consider again Lowe's (2003) model of mental causation:



Lowe insisted on the fact that the exercise of M 's causal powers is *invisible*.²¹ On the one hand, it is somewhat dubious that the tools that are available to scientists are able to detect causes other than the physical ones. On the other hand, since M and P^* are instantiated at the same moment t , any external observer is likely to conclude that P^* is *immediately* sufficient for the production of P_1 . After all, empirical data reveal that the instantiation of P^* at t is systematically followed by the instantiation of P_1 at t_1 . The possibility of a non-physical causal intermediary would not even be considered.²²

²⁰Montero argues that 'while we certainly do not have a complete nonmental account of what we take to be mental causes, we have a good start' (2003, p. 185).

²¹This is ultimately the reason why Montero (2003) is skeptical about the argument from physiology.

²²Robb (2018) has argued against the invisibility claim. He contends that there is, at least in principle, a way to empirically determine whether strongly emergent mental properties have a

Interestingly, Lowe's model need not be actual to represent an obstacle for the argument from physiology. Even the mere possibility that mental causation is invisible threatens the efficacy of the first component of the argument, since the physicalist cannot exclude that such a possibility actually obtains (Owen 2020). Accepting that mental causation *could* be invisible is enough for undermining the inference from the absence of evidence to the evidence of absence the physicalist relies on.²³

That said, let us suppose that either Lowe's model is implausible, or that Lowe is wrong about the invisibility of mental causation. Still, there might be some tension when it comes to the claim that 'there is no principled *a priori* reason why 20th-century physiological research should not have uncovered special mental and vital forces' (Montero and Papineau 2016). A potential problem stems from the ambiguous status of the causal closure principle. Here, CC has been presented as a substantial, metaphysical claim about the causal structure of the universe. However, its *methodological* implications should be made clear as well. In fact, as Kim (1996, pp. 147–148) points out, many physicalists 'accept the causal closure of the physical not only as a fundamental metaphysical doctrine but as an indispensable methodological presupposition of the physical sciences'.²⁴

The claim that the causal closure principle also has a methodological component seems to be a plausible one. Focusing on physiology and neuroscience, the following methodological precept seems to drive current research:

(MCC) when accounting for the production of human behaviour, the *explanans* cannot include causes other than physical ones²⁵

However, this might have an impact on the physicalist's line of reasoning. With MCC in the picture, the physicalist may be prevented from legitimately resorting to the first component of the argument from physiology. Assume that *sui generis*

causal role – more on this in Sect. 5. However, even if Robb is right, mental causation *à la* Lowe is nonetheless *almost* invisible. This is enough for undermining the first component of the argument from physiology.

²³ An anonymous reviewer cast doubts on this point by providing a counterexample: the absence of evidence that there was a tornado an hour ago is good evidence that there was no tornado, even if one cannot exclude the possibility that there was actually a tornado but the damages were immediately and silently repaired. As far as I can see, there is a crucial difference between this scenario and the case of invisible mental causation. We know that tornados can leave evidence in the actual world – and they usually do, which explains why we regard the described scenario as an unlikely one. Provided that we could be wrong, the fact that we have repeatedly observed such evidence allows us to legitimately infer that there was no tornado an hour ago. On the contrary, we do not know whether invisible mental causation takes place in the actual world. This makes the inference from the absence of evidence to the evidence of absence much riskier.

²⁴ See also Zargar et al. (2020).

²⁵ Admittedly, MCC is the methodological version of SCC, that is not the principle I am considering. To be consistent, I should take MCC to be the precept that when accounting for the production of behaviour, the *explanans* cannot include causes other than *immediately sufficient* physical ones. As far as I can see, the simplification does not affect in any significant way the considerations made here.

mental properties actually do have a causal role in the production of behaviour. Suppose then that we have a somewhat fine-grained account of what goes on in the brain between, say, the second before a subject's decision to move their left arm and the contraction of their muscles. Even if some gaps were detected in the physical causal chain, neuroscientists would not even think about *sui generis* mental causation. In accordance with MCC, they would simply keep searching for some physical events that may have gone unobserved. *Contra* Montero and Papineau, if scientific investigation in neuroscience is driven by something along the lines of MCC, there seem to be principled reasons to doubt that recent research could have uncovered special mental forces. Even if we had an account of the physical causal processes within the brain that is detailed enough to allow for the detection of causal gaps, such negative evidence about the possibility of non-physical forces operating would simply be mistaken for ignorance about the existence of further physical causes.

At this stage, Montero and Papineau could insist that the causal closure principle has its methodological implications *in virtue of the fact* that it is a strongly empirically supported metaphysical claim. Put another way, scientists would exclusively look for physical causes because *prior* evidence has suggested that physical events have only physical causes. This sounds plausible, although exploring the issue would require a detailed historical analysis. In general, however, it is fair to say that physicalists should be careful when arguing along with Kim that CC has strong methodological implications. More specifically, they have to point out that the metaphysical component is prior – historically and epistemically – to the methodological one. Otherwise, one may suspect that the valid negative evidence for CC is not as much as we tend to think.

4.2 *The Second Component*

At this stage, let us consider the second component of the argument from physiology, that focuses on the *positive* evidence we have available. In particular, Papineau (2001, p. 31) insists on the development of biology and neuroscience in the twentieth century:

[...] the catalytic role and protein constitution of enzymes were recognized, basic biochemical cycles were identified, and the structure of proteins analyzed, culminating in the discovery of DNA. In the same period, neurophysiological research mapped the body's neuronal network and analysed the electrical mechanisms responsible for neuronal activity.

Briefly, the idea seems to be the following: the more we learn about causal processes in living organisms without finding evidence for *sui generis* mental causation, the less room for *sui generis* mental causation is left. Note that this is not the same as arguing that we have not found any trace of *sui generis* mental causation. What the physicalist insists on, in this case, is that recent research has identified a number of physical causes of our behaviour. The strength of this line of reasoning directly depends on our positive knowledge of the brain's functioning. If we can count on a

sufficiently precise account of the physical processes that are responsible for the production of our behaviour, we may have reasons to think that *sui generis* mental properties do not have a role.

Unfortunately for the physicalist, optimism in the actual explanatory power of neuroscience and physiology may be misplaced. To be clear, I am not adopting an antiscientific stance. On the contrary, I assume that philosophy cannot ignore the empirical evidence we have available. Narrowing it down to the mind, philosophers should not overlook the results that neuroscience and psychobiology have been providing. At the same time, I am not calling into question the import of the advancements that have been made in these disciplines. New tools have opened new possibilities for studying the brain, both structurally and functionally, and our understanding of the physiology and organisation of our nervous system has dramatically improved in recent years. The point is just that unrealistically high expectations seem to be put on the explanatory power of current science.

Coming back to the first component of the argument from physiology, let us consider once again the example of ghosts' existence, that Montero (2003) takes to be a case in which the absence of evidence counts as evidence of absence. As she points out, the sine qua non condition for this to be possible is that we have a 'pretty good understanding of what actually causes those bumps in the night that scare people into thinking their houses are haunted' (2003, p. 185). Accordingly, when we consider the absence of evidence for *sui generis* mental causation and we aim at inferring that there are no special mental forces operating, we need a 'pretty good understanding' of what actually causes our behaviour. Needless to say, it is not enough to know that the activation of specific brain areas has a causal role in the production of certain responses. This coarse-grained knowledge is perfectly compatible with the exercise of non-physical causal powers somewhere in the causal chain. What seems to be needed as background knowledge is a reasonably good approximation of a *complete* and *fine-grained* description of the causal processes that are supposed to produce our behaviour. If this description involves no *sui generis* mental causes, then we may have good reasons to believe in the absence of non-physical causes of behaviour.

A similar level of detail seems to be required in the second component of the argument from physiology. The more psychophysiology and neuroscience are successful in providing detailed, purely physical descriptions of the causal processes that bring about our behaviour, the less *sui generis* mental causes fit in the picture – here, the assumption is that *sui generis* mental causation would have effects on the brain. The fewer links in the causal chain are left unidentified, the more *sui generis* mental causation turns out to be implausible. Again, what the physicalist needs to presuppose to rule out special mental forces seems to be that psychophysiology and neuroscience have already provided a reasonably fine-grained and almost complete account of the way our behaviour is produced by physical processes.

Unfortunately, this confidence may not be justified. Let us focus, along with Owen (2020), on a passage from a recent book by Christof Koch, one of the most prominent neuroscientists of consciousness. Referring to our 'inadequate knowledge

of the prodigious complexity of the brain, from the molecular to the system level', Koch (2019, p. 138) writes:

The dirty secret of computational neuroscience is that we still do not have a complete dynamic model of the nervous system of the worm *C. elegans*, though it only has 302 nerve cells and its wiring diagram, its connectome, is known. So here we are, trying to understand the human brain, when we do not yet understand the worm brain.²⁶

This may seriously undermine the soundness of the physicalist's line of reasoning. On the one hand, the absence of evidence concerning *sui generis* mental causation cannot count as evidence of absence, since we lack that 'pretty good understanding' of the brain's functioning that would make the inference legitimate. On the other hand, since our knowledge of the causal interactions within the brain is far from being complete, we cannot exclude the possibility of non-physical causes intervening at some stage of the causal process.

Taking stock, both the components of the argument from physiology seem to rely on an excessively optimistic assumption concerning our knowledge about the functioning of the nervous system, which is arguably what would be influenced by *sui generis* mental causes. At this stage, physicalists may reply that we need not assume that we have a fine-grained and complete account of the causal processes within the brain. What we need, they may argue, is a reasonably detailed description of the functioning of the nervous system's working units.

In a recent discussion of the causal argument, Papineau (2020, p. 16) explicitly refers to such knowledge as the empirical ground for the argument from physiology:

It was only in the middle of the twentieth century that a detailed understanding of the electro-chemical workings of neurons convinced the scientific mainstream that there is no place for *sui generis* mental forces.

The point is that we know which kinds of neurotransmitters, receptors, and molecules are involved in synaptic transmission, and we can count on a reasonably precise reconstruction of the way they causally interact. Not by chance, the process of synaptic transmission is often used in the philosophy of science as an example of a phenomenon that was accounted for in mechanistic terms (among others, see Craver 2007). Whether this kind of evidence is sufficient to vindicate CC is the question I will address in the remainder of this article.

5 Synaptic Transmission, Causal Closure and Emergence

When it comes to the mechanisms of synaptic transmission, a sufficiently detailed story of the involved causal processes seems to be available. Given two neurons n_1 and n_2 , the electrical impulse reaching the synaptic vesicles at the end of n_1 's axon arguably counts as the immediate, sufficient physical cause of neurotransmitters

²⁶See also Garcia (2014); Di Francesco and Tomasetta (2015).

being released in the synaptic cleft between n_1 and n_2 . In its turn, the release of neurotransmitters in the cleft seems to be the immediate, sufficient physical cause of the activation of specific receptors on the post-synaptic membrane of n_2 . Ultimately, the result is that n_2 is either excited or inhibited. Regardless of the oversimplification, the metaphysical take-home message is clear. At least *prima facie*, there seems to be evidence for what I will refer to as the *synaptic* causal closure thesis:

(Synaptic-CC) physical events *within the synaptic micro-system* have immediate and sufficient physical causes

True, one may argue that we cannot be *sure* that we know everything about the causal steps involved in the mechanisms in question. Hence, there could be room for the intervention of non-physical properties even in the processes of synaptic transmission. In what follows, however, I will just *assume* that the electrochemical properties physical sciences are familiar with are sufficient – and, more precisely, *immediately* sufficient – to account for neurotransmission. Still, I contend that something more seems to be needed if the ultimate purpose of the physicalist is to exclude the possibility of interactionist dualism.

Upon closer inspection, the fact that *sui generis* mental properties are not involved in the process of synaptic transmission is perfectly compatible with the claim that they do play a role in the production of our behaviour. As the story goes, dualists argue that *sui generis* mental properties emerge when a certain level of complexity is reached within the nervous system. Accordingly, they are not compelled to claim that *sui generis* mental properties exercise their powers within the processes of synaptic transmission, that serve as building blocks for the functioning of the whole. Such properties may well exercise their powers within the nervous system at a higher level of organisation.

That said, physicalists are arguably aware of the problem. When focusing on the mechanisms of synaptic transmission, they seem to resort to an implicit further premise, namely that the behaviour of the nervous system is *compositionally determined* by the behaviour of its working units, that are neurotransmission mechanisms.²⁷ In other terms, the behaviour of the nervous system would be *completely* determined by (i) the behaviour of its components and (ii) the way these components are spatiotemporally organised. Note that this is not the same as saying that the behaviour of the nervous system's organised components is *nomologically* sufficient for the behaviour of the whole. This would leave room for the non-overdetermining contribution of *sui generis* emergent mental properties, that are usually taken to be nomologically dependent on the physical configurations they emerge from. If an analysis in terms of nomological sufficiency were to be provided, then the physicalist should make explicit that the involved natural laws are physical ones. In this case, the transitivity of nomological sufficiency would be blocked. Physical laws do not

²⁷I am grateful to David Papineau for pointing out this to me.

account for the emergence of *sui generis* mental properties, that is governed by *special* natural laws.²⁸

Avoiding unnecessary complications, the point is that the physicalist takes the behaviour of the nervous system to be the sum of the behaviours of the neurons – and glial cells, and so forth – it is made up of, and nothing else. Clearly, once this premise is brought into the picture, we can easily infer CC. If we are justified in holding Synaptic-CC and we have reasons to think that all causal processes within the nervous system are nothing but sums of mechanisms of synaptic transmission, then it follows that we are justified in believing that all physical events within the nervous system have immediately sufficient physical causes.

Framing the discussion in terms of combination principles can be useful. I have conceded that the physicalist can rely on a fairly detailed understanding of the physical causes involved in neurotransmission processes. However, this understanding would be useless without some insights into the way these processes combine. In order to take the evidence for Synaptic-CC to be evidence for CC as well, the physicalist takes for granted an *additive* principle of composition concerning the internal causal organisation of the nervous system:

(Additivity) powers in combination produce the sum of the manifestations they produce independently²⁹

Some specifications are necessary. Consider the case of a watch. When compared to the nervous system, it turns out to be a fairly simple mechanism. Even in such a case, however, it would be trivial to point out that the whole does something that its components, as well as the mere sum of them, cannot do. A hand alone does not tell the time, nor does a heap of gears randomly combined. Clearly, the components of the watch have to be spatiotemporally organised in an appropriate fashion. This is arguably valid for the greatest majority of the systems one could consider, even for the least complex ones. In what follows, I will not further discuss this issue, and I will take the appropriate-organisation clause to be implicit in Additivity.

At this stage, unsurprisingly, it all comes down to the following question: when it comes to the nervous system, is there evidence – be it theoretical or empirical – for Additivity or some similar combination principles that allow for the inference from Synaptic-CC to CC? If not, then the argument from physiology fails to vindicate a version of the causal closure principle that is strong enough to rule out the possibility of causally efficacious emergent mental properties.

Before concluding by discussing a possible strategy the physicalist could resort to, let me briefly make a related point. If one thinks about it, Additivity is exactly what strongly emergent properties violate (Robb 2018). Let us consider the characterisation of emergent properties as properties that ‘confer causal capacities on the object that go beyond the summation of capacities directly conferred by the object’s microstructure’ (O’Connor and Wong 2005, p. 665). Now, suppose that Additivity

²⁸On this point, see Yates (2009).

²⁹I borrow this formulation from Robb (2018). Note that Robb is not committed to the principle.

was somehow vindicated with respect to the nervous system. The physicalist would have *direct* reasons against strong emergentism about mental properties, and it is unclear whether the causal argument would still have a role in the dialectic between the physicalist and the interactionist dualist. In other words, there seems to be a sense in which evidence for Additivity would be crucial in the economy of the causal argument, since it would allow for the inference from Synaptic-CC to CC. However, this evidence would also make the causal argument somewhat redundant. Were Additivity vindicated, interactionist dualism would be already out of the picture.

That said, let us leave this issue aside and focus on a possible way to support Additivity. Vindicating the claim that the internal causal structure of the nervous system is additive is far from being an easy task. As far as I can see, the best way to do it requires two steps:

1. identifying the causal profiles of the individual working units of the nervous system and the physical laws that govern their interactions;
2. once we have *complete* knowledge about them, checking whether the occurrence of the physical effects that are usually supposed to have mental causes follows.

If the answer to (2) is positive, then Additivity – and therefore CC – is vindicated. Crucially, this is not the same as proving that *sui generis* mental properties have no causal power at all. As we have seen, the possibility of overdetermination is compatible with CC holding.

Unfortunately for the physicalist, even if the problem of the calculation's difficulties is left aside, this strategy is hardly viable. In particular, a couple of points are worth highlighting. Admittedly, it is difficult to deny that, if our behavioural responses follow from the summation of the (organised) powers of the physical micro-components of the nervous system, then *sui generis* mental properties could be either causally inefficacious or at best genuinely overdetermining. However, things get complicated if (2) is answered negatively. In principle, this could count as evidence for the emergentist view that *sui generis* mental properties play an ineliminable role in the production of our behaviour. Physicalists, however, would arguably resist such a conclusion. More likely, they would argue that we have simply failed to identify some physical micro-powers or laws. Note that, in doing so, the physicalist would not be stubbornly begging the question. Indeed, (2) requires that we have *complete* knowledge of the powers of the nervous system's working units and the relevant physical laws governing their interactions. However, it is far from clear that we can come to know, at some point, that our knowledge about them is complete.

So far so good, at least for the physicalist. If the answer to (2) is positive, Additivity is vindicated. If the answer is negative, Additivity is still a possibility, although not an empirically supported one. The real issue with the strategy in question is that, once again, it is extremely demanding in terms of knowledge of the microphysical processes within our nervous system. Our brain is an incredibly complex object, to say the least, made up of 86 billion neurons connected by a huge number of synapses. Clearly, it is not the case that our brain is *entirely* responsible for each of our behavioural responses. Still, the size and the intricacy of the nervous

activations that are supposed to be causally responsible for our behaviours make (1) an utterly unrealistic goal to achieve, at least given the current state of research and the tools we have available. True, the study of the physiology and the functioning of the brain has made great progress in the last decades. Still, we are far from knowing the precise causal role of all the individual micro-components of the nervous system, and we largely ignore how they work together when it comes to bringing about our behaviour. Maybe, in the future, neurophysiology will provide detailed, microscopical descriptions of the causal processes taking place within our nervous system. Currently, however, this is nothing more than an (extremely) optimistic expectation concerning the development of brain science. Unfortunately for the physicalist, this is not enough for vindicating Additivity.

Taking stock, the physicalist's move to insist on the mechanisms of synaptic transmission is promising, at least *prima facie*. As we have seen, there is a considerable amount of empirical evidence suggesting that neurotransmission mechanisms involve only physical causes that are immediately sufficient for their effects. The problem is that, to legitimately generalise this up to CC, the physicalist should be in the position to take for granted something along the lines of Additivity with respect to the nervous system. Vindicating the claim that the micro-powers of the nervous system combine in an additive fashion, however, is not an easy task. I have briefly discussed what seems to be the most straightforward way to do it. However, resorting to the two-step strategy I have outlined is clearly not an option, at least at the moment.

Acknowledgments I am grateful to David Papineau, Alfredo Tomasetta and Michele Di Francesco for the numerous discussions on the causal argument. They have also provided me with extremely helpful comments and feedback on various drafts of this work. An earlier version of this paper was presented at the tenth edition of the Research Workshop on Philosophy of Biology and Cognitive Science (PBCS X). I would like to thank the audience for their questions and suggestions. Finally, I am grateful to two anonymous reviewers for providing extensive and helpful comments.

References

- Alter T (2021) A defense of the supervenience requirement on physicalism. *Thought: A J Philos* 10(4):264–274
- Baumeister RF, Lau S, Maranges HM, Clark CJ (2018) On the necessity of consciousness for sophisticated human action. *Front Psychol* 9:1925
- Baysan U (2020) Causal emergence and epiphenomenal emergence. *Erkenntnis* 85(4):891–904
- Bennett K (2003) Why the exclusion problem seems intractable, and how, just maybe, to tract it. *Noûs* 37(3):471–497
- Bernstein S (2016) Overdetermination underdetermined. *Erkenntnis* 81(1):17–40
- Bruntrup G, Jaskolla L (eds) (2016) *Panpsychism: contemporary perspectives*. Oxford University Press, New York
- Campbell K (1970) *Body and mind*. Macmillan, London; Toronto
- Chalmers DJ (1996) *The conscious mind: In search of a fundamental theory*. Oxford University Press, Oxford

- Chalmers DJ (2006) Strong and weak emergence. In: Clayton P, Davies P (eds) *The re-emergence of emergence*. Oxford University Press, New York, pp 244–255
- Chalmers DJ (2010) *The character of consciousness*. Oxford University Press, New York
- Crane T, Mellor DH (1990) There is no question of physicalism. *Mind* 99(394):185–206
- Craver CF (2007) *Explaining the brain: mechanisms and the mosaic Unity of neuroscience*. Oxford University Press, New York
- Dasgupta S (2014) The possibility of physicalism. *J Philos* 111(9/10):557–592
- De Brigard F (2014) Self-stultification objection. *J Conscious Stud* 21(5–6):120–130
- Dennett DC (1991) *Consciousness explained*. Little, Brown and Company, Boston
- Di Francesco M, Tomasetta A (2015) The end of the world? Mental causation, explanation and metaphysics. *Humana Mentis* 8(29):167–190
- Dimitrijević DR (2020) Causal closure of the physical, mental causation, and physics. *Eur J Philos Sci* 10:1
- Fodor JA (1989) Making mind matter more. *Philos Top* 17(1):59–79
- Garcia R (2014) Closing in on causal closure. *J Conscious Stud* 21(1–2):96–109
- Gibb S (2010) Closure principles and the laws of conservation of energy and momentum. *Dialectica* 64(3):363–384
- Gibb S (2015) The causal closure principle. *Philos Q* 65(261):626–647
- Goff P (2017) *Consciousness and fundamental reality*. Oxford University Press, New York
- Gomes G (2002) The interpretation of Libet's results on the timing of conscious events: a commentary. *Conscious Cogn* 11(2):221–230
- Jackson F (1982) Epiphenomenal Qualia. *Philos Q* 32(127):127–136
- James W (1879) Are we automata? *Mind* 4:1–22
- Kim J (1976) Events as property exemplifications. In: Brand M, Walton D (eds) *Action theory*. Springer, Dordrecht, pp 159–177
- Kim J (1996) *Philosophy of mind*. Westview Press, Boulder
- Kim J (1998) *Mind in a physical world. An essay on the mind-body problem*. MIT Press, Cambridge, MA
- Kim J (1999) Making sense of emergence. *Philos Stud* 95:3–36
- Kim J (2005) *Physicalism, or something near enough*. Princeton University Press, Princeton; Oxford
- Koch C (2019) *The feeling of life itself: why consciousness is widespread but can't be computed*. MIT Press, Cambridge
- Lavazza A (2016) Free will and neuroscience: from explaining freedom away to new ways of operationalizing and measuring it. *Front Hum Neurosci* 10:262
- Lewis D (1983) New work for a theory of universals. *Australas J Philos* 61(4):343–377
- Libet B (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav Brain Sci* 8(4):529–566
- Lowe EJ (2000) Causal closure principles and emergentism. *Philosophy* 75(294):571–585
- Lowe EJ (2003) Physical causal closure and the invisibility of mental causation. In: Walter S, Heckmann H (eds) *Physicalism and mental causation: the metaphysics of mind and action*. Imprint Academic, Exeter, pp 58–78
- McLaughlin BP (1992) The rise and fall of British emergentism. In: Beckerman A, Flohr H, Kim J (eds) *Emergence or reduction? Essays on the prospects of non-reductive physicalism*. De Gruyter, Berlin, pp 49–93
- Mele AR (2014) *Free: why science hasn't disproved free will*. Oxford University Press, New York
- Melnyk A (2006) Realization and the formulation of physicalism. *Philos Stud* 131(1):127–155
- Montero BG (2003) Varieties of causal closure. In: Walter S, Heckmann H (eds) *Physicalism and mental causation: the metaphysics of mind and action*. Imprint Academic, Exeter, pp 173–187
- Montero BG (2013) Must physicalism imply the supervenience of the mental on the physical? *J Philos* 110(2):93–110
- Montero BG, Brown C (2018) Making room for a this-worldly physicalism. *Topoi* 37(3):523–532

- Montero BG, Papineau D (2005) A defence of the *via negativa* argument for physicalism. *Analysis* 65(3):233–237
- Montero BG, Papineau D (2016) Naturalism and physicalism. In: Clark KJ (ed) *The Blackwell companion to naturalism*. Wiley, Malden and Oxford, pp 182–195
- Noordhof P (2010) Emergent causation and property causation. In: Macdonald C, Macdonald GF (eds) *Emergence in mind*. Oxford University Press, New York, pp 69–99
- O'Connor T (1994) Emergent properties. *Am Philos Q* 31(2):91–104
- O'Connor T, Wong HY (2005) The metaphysics of emergence. *Noûs* 39(4):658–678
- Owen M (2020) The causal efficacy of consciousness. *Entropy* 22(8):823
- Papineau D (1998) Mind the gap. *Philos Perspect* 12:373–388
- Papineau D (2001) The rise of physicalism. In: Gillett C, Lower B (eds) *Physicalism and its discontents*. Cambridge University Press, Cambridge, pp 3–36
- Papineau D (2002) *Thinking about consciousness*. Oxford University Press, New York
- Papineau D (2009) The causal closure of the physical and naturalism. In: Beckermann A, McLaughlin BP, Walter S (eds) *The Oxford handbook of philosophy of mind*. Oxford University Press, New York, pp 53–65
- Papineau D (2020) The problem of consciousness. In: Kriegel U (ed) *The Oxford handbook of the philosophy of consciousness*. Oxford University Press, New York, pp 14–38
- Pereboom D (2011) *Consciousness and the prospects of physicalism*. Oxford University Press, Oxford
- Pockett S, Purdy SC (2011) Are voluntary movements initiated preconsciously? The relationship between readiness potentials, urges and decisions. In: Sinnott-Armstrong W, Nadel L (eds) *Conscious will and responsibility*. Oxford University Press, New York, pp 34–46
- Popper KL, Eccles JC (1977) *The self and its brain*. Springer, Berlin
- Robb D (2018) Could mental causation be invisible? In: Carruth A, Gibb S, Heil J (eds) *Ontology, modality and mind: themes from the metaphysics of E. J. Lowe*. Oxford University Press, New York, pp 165–176
- Robinson WS (2019) *Epiphenomenal mind: an integrated outlook on sensations, beliefs, and pleasure*. Routledge, New York; London
- Shapiro L (ed) (2007) *The correspondence between Princess Elisabeth of Bohemia and Descartes*. The University of Chicago Press, Chicago and London
- Sider T (2003) What's so bad about overdetermination. *Philos Phenomenol Res* 67(3):719–726
- Tomasetta A (2015) Physicalist naturalism in the philosophy of mind. *Discipline Filosofiche* 25(1): 89–111
- Wilson J (2015) Metaphysical emergence: weak and strong. In: Bigaj T, Wüthrich C (eds) *Metaphysics in contemporary physics*. Brill, Leiden and Boston, pp 345–402
- Wilson J (2021) *Metaphysical emergence*. Oxford University Press, New York
- Yates D (2009) Emergence, downwards causation and the completeness of physics. *Philos Q* 59(234):110–131
- Zargar Z, Azadegan E, Nabavi L (2020) Should methodological naturalists commit to metaphysical naturalism? *J Gen Philos Sci* 51:185–193