CHAPTER 1

Physicalism

In this chapter I introduce the first element of the physicalist triad. Physicalism is the view that everything that exists is either itself physical or something composed of, constituted by, realised by, grounded in, or in some other way 'nothing over and above' physical entities. Physicalism is a popular account of the mind. Very roughly, physicalism about the mind says that mental entities (states, properties, events, processes) are nothing over and above physical entities, which are brain states or brain activity on some versions of physicalism. The strongest argument for physicalism takes the existence of mental causation as a premise, and concludes that mental causation can only be possible, given certain assumptions about what causation in the actual world must be like, if some form of physicalism is true. Versions of this argument include the 'causal argument for physicalism' championed by David Papineau (1993; 2001; 2002), Donald Davidson's (1970) argument for anomalous monism, and Jaegwon Kim's (1993; 1998; 2001) 'causal exclusion argument'.

In what follows, I outline these arguments in detail. My aim in this chapter is not to challenge these arguments by disputing their premises. Instead, my aim is to reveal some important assumptions that are implicit in these arguments. The most important of these is an assumption of what mental causation is. Arguments for physicalism assume that mental causation is mental items (events, processes or states) standing in causal relations to physical items (e.g. movements of a person's body). I call this the relational understanding of mental causation. I aim to show in Section 1.2 that this assumption is essential to arguments for physicalism: without it the arguments are invalid. I also argue, in Sections 1.3 and 1.4, that the plausibility of the relational understanding of mental causation in turn derives from two other philosophical theories: causal theories of intentional action and relational approaches to causation. According to causal theories of intentional action, intentional or voluntary human action is possible only if mental items stand in causal relations to physical events such as bodily movements. According to relational approaches to

How to cite this book chapter:
White, Andrea. 2024. *Understanding Mental Causation*. Pp. 13–37. York: White Rose University Press. DOI: https://doi.org/10.22599/White.b.

causation, any naturalistic account of causation is committed to the idea that all causation is the same; causation is a homogenous phenomenon. Causation can hold between diverse relata, but there is one relation of causation, and that is all causation is. On this approach, what makes an example of causation 'mental' can only be that one or both of the causally related entities is mental.

In the last section of this chapter, Section 1.5, I outline reasons why someone might be dissatisfied with physicalism as a metaphysics of mind. I do not intend to disprove physicalism in this section. I only intend to show that physicalism is not without its critics. The causal arguments for physicalism that I describe in Section 1.1 can make it seem like physicalism is the only option. Showing that there are some reasons to doubt the truth of physicalism helps justify my project, which is to show that the causal arguments for physicalism are artificially bolstered by their association with causal theories of intentional action and relational approaches to causation.

1.1 Physicalism and mental causation

In philosophy of mind, the problem of mental causation is the problem of how that which is mental is able to causally interact with that which is physical. This problem is usually presented as a 'how possibly' question, that is, a question about how mental causation could exist. It is usually accepted as prima facie true that mental causation does exist; the difficulties arise only when we try to understand how this could be, given certain assumptions about the fundamental nature of reality. As Peter Menzies states, 'philosophical questions about mental causation revolve around … how it is possible in the first place in the light of certain metaphysical assumptions and principles' (2013: 58). The metaphysical assumptions and principles that seem to make the existence of mental causation puzzling concern the apparent physicality of the causal world. For example, Kim tells us that the problem of mental causation is 'to explain how mentality can have a causal role in a world that is fundamentally physical' (2005: 1). The existence of mental causation is thought to be especially difficult to reconcile with the principle of the causal closure of the physical world, which says that 'at every time at which a physical effect has a cause it has a sufficient physical cause' (Gibb 2013: 2).[4]

[4] There are many alternative formulations of the principle of causal closure, for example: 'any physical state or change, if it has a cause or explanation, has a physical cause or explanation' (Hopkins 1978: 223); '[f]or all physical events and states there are necessary and sufficient physical conditions, their "explanations" or "causes"' (Skillen 1984: 514); '[e]very physical effect has its chance fully determined by physical events alone' (Noordhof 1999: 367); and '[a]ll physical effects are fully determined by law by prior physical occurrences' (Papineau 2001: 9). These various formulations are

Proposing a metaphysics of mind that reconciles the existence of mental causation with principles about what causation in the actual world must be like, such as the principle of causal closure, has been the aim of a substantial amount of research in philosophy of mind. A key aim in contemporary philosophy of mind is to establish a metaphysics of mind that (a) saves the phenomenon of mental causation and (b) respects plausible principles about what actual causation is like, such as the principle of causal closure. Physicalism, of some form or another, is generally considered to be the metaphysics that has the best chance of satisfying these two conditions.

Physicalism says that everything that exists is either itself a physical entity or 'nothing over and above' a physical entity. Exactly what it is for one entity to be 'nothing over and above' another is a matter of debate. Some physicalists express their position in terms of *supervenience* (Haugeland 1982; Hellman and Thompson 1975). This version of physicalism says that all the non-physical aspects of the world supervene on the physical aspects, which is to say that it is impossible for two things to differ with respect to their non-physical properties without also differing in their physical properties. There are a number of difficulties associated with expressing physicalism in terms of supervenience.

First, supervenience can be global or local. Physicalism expressed in terms of global supervenience says that no two *worlds* can differ with respect to their non-physical properties without also differing in their physical properties. As Frank Jackson puts it, '[a]ny world which is a minimal physical duplicate of our world is a duplicate of our world' (1998: 14). Physicalism expressed in terms of local supervenience says that no *person/event/object* can differ with respect to their non-physical properties without also differing in their physical properties. For example, Davidson claims that supervenience of the mental on the physical amounts to the claim that 'there cannot be two events alike in all physical respects but different in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect' (1970/2001: 214).

One reason to opt for global supervenience rather than local supervenience is that global supervenience accommodates externalism about the content of mental states. Externalism holds that the content of mental states is determined in part by the nature of the environment and not by the intrinsic properties of the thinker. For example, suppose Oscar thinks 'water is wet'. Externalism says that, for Oscar's thought to be the thought that it is, Oscar must be related to

---

not equivalent. Furthermore, the principle of causal closure is supposed to be derived from, and supported by, the findings of scientific investigations into causal processes, but it is a contentious question which, if any, of the various formulations of causal closure enjoy such support. See Lowe (2000) for a discussion of issues relating to the principle of causal closure, and see Papineau (2001) for a defence of the principle.

certain external states of affairs. For Oscar's thought to be about Earth-water, the stuff occupying rivers that is composed of hydrogen and oxygen, Oscar's environment must contain Earth-water or his linguistic community must have been in contact with Earth-water. Twin-Oscar, who lives in an alternative world that does not contain Earth-water but instead contains a substance with exactly the same observable properties but which has a different molecular structure, cannot have the same thoughts as Oscar. This is so even if both Oscar and Twin-Oscar know nothing about the molecular structure of the watery stuff that exists in their worlds and are disposed to ascribe the exact same properties to the watery stuff of their worlds. Externalism is a plausible and well supported thesis (Burge 1979; Putnam 1975) so it would be good for physicalism to be consistent with it. However, the difficulty with expressing physicalism in terms of global supervenience is that the theory does not seem to be meaningfully different to non-physicalist views of the mind such as emergentism, the view that the mental aspects of the world depend on the physical aspects of the world but have their own independent causal powers (thus causal closure is false) (Horgan 1993; Wilson 2005). Expressing physicalism in terms of local supervenience makes physicalism a more substantive thesis, but one that has a higher chance of being false. Externalism about mental contents seems to disprove it: even if we assume that Oscar and Twin-Oscar have all the same intrinsic physical properties, the fact that they live in different worlds entails that they have different thoughts.

Another issue with expressing physicalism in terms of supervenience is that supervenience captures covariance between supervenient and subvening properties but that is all. Therefore, to claim that everything that exists supervenes on the physical is not very informative. To say that mental properties supervene on physical properties does not tell us very much about how the mental and the physical relate. If one property can be shown to be identical to another, then there will be a relation of supervenience between them. However, supervenience also holds between properties related by realisation, constitution, or a determinable–determinate relation. What these relations have in common is that they are all one-way necessitation relations: the subvening property necessitates the presence of the supervenient property but not the other way around. As Terence Horgan argues, if physicalists express their view in terms of supervenience they must also explain why supervenience holds between the physical and the non-physical in a 'materialistically acceptable way' (1993: 556). In other words, physicalists need to elucidate the relation between the physical things and the putative non-physical things that explains why supervenience holds, and why the physical things are more fundamental or more real or, in some other way, ontologically superior. As Amanda Bryant puts it, 'supervenience formulations of physicalism fail to capture or illuminate the common physicalist contention that the mental metaphysically depends on the physical' (2020: 489). In response to this challenge, some physicalists have expressed their view in terms of realisation: non-physical properties are

realised by physical properties (Melnyk 2003; Melnyk 2006; Melnyk 2018; Shoemaker 2001; Shoemaker 2007); others in terms of constitution (or an analogue of constitution): non-physical properties are constituted by physical properties (Pettit 1993); others in terms of a determinable–determinate relation: non-physical properties are determinable properties of which physical properties are the determinates (Yablo 1992); and others in terms of fixing: non-physical properties fix all other properties (Elpidorou & Dove 2018); truthmaking: physical facts make true all facts (Morris 2018); and, more recently, grounding: non-physical properties and facts are grounded by physical properties and facts (Bryant 2020). It would take us too far afield to assess each of these proposals individually. The debate concerning what it means to say that everything that exists is nothing over and above the physical is ongoing, with some writers more pessimistic than others about whether the physicalist can adequately meet this challenge (examples of pessimists include Horgan (1993), Lynch and Glasgow (2003) and Wilson (2016)).

Regardless of how the nothing-over-and-above relation should be spelt out, several arguments conclude that some form of physicalism about mentality must be true, because otherwise an account of mental causation that respects plausible principles about what causation is like (such as causal closure) is impossible. Three such arguments are the 'causal argument for physicalism' championed by Papineau (1993; 2001; 2002),[5] Davidson's (1970) argument for anomalous monism, and Kim's (1993; 1998; 2001) 'causal exclusion argument'.

As Papineau (2001: 9) presents it, the causal argument for physicalism has three premises:

1. 'All mental occurrences have physical effects'.
2. 'All physical effects are fully determined by law by prior physical occurrences' (this is Papineau's formulation of the causal closure principle).
3. 'The physical effects of mental causes are not all overdetermined'.

From these three premises, it is concluded that '[m]ental occurrences must be identical with physical occurrences' (2001: 9). The causal argument is used to establish a physicalist identity theory about mental *occurrences* or *events*. However, proponents of the causal argument typically assume that events, processes and states are not significantly different in nature, and so what goes for mental *events* goes for mental *processes* and mental *states* (or at least 'token states') as well. Importantly, the first premise is understood by proponents of the causal argument as simply equivalent to the claim that mental causation exists.

Davidson's argument for anomalous monism is another example of an argument for identifying mental events with physical events that takes the

[5] This argument is also known as 'the causal overdetermination argument' (Crane 1995; Gibb 2013) and 'the overdetermination argument' (Noordhof 1999; Sturgeon 1998).

existence of mental causation as a premise. Anomalous monism asserts that every individual mental event is identical with some physical event, but mental kinds are distinct from physical kinds. So, for example, even though *being a decision* cannot be identified with any physical kind, not even *being a brain event*, every token decision is identical with a physical event of some kind. Davidson's (1970/2001: 208) argument for this view involves three premises:

1. 'At least some mental events interact causally with physical events' (the 'principle of causal interaction').
2. 'Where there is causality, there must be a law: events related as cause and effect fall under strict deterministic laws' (the 'principle of the nomological character of causality').
3. 'There are no strict deterministic laws on the basis of which mental events can be predicted and explained' (the 'anomalism of the mental').

Again, the first premise of this argument is typically understood as an assertion that mental causation exists.

Kim objects to Davidson's anomalous monism on the grounds that 'on anomalous monism, events are causes or effects only as they instantiate physical laws, and this means that an event's mental properties make no causal difference' (1989: 34–35). On Kim's view, events are causes in virtue of the properties they involve, and not every property an event involves causally matters, unless there is causal overdetermination. Kim argues that the nomological character of causality and anomalism of the mental imply that it is always an event's *physical* nature that is causally relevant; whatever *mental* properties an event may involve are excluded from being causally relevant by physical properties, which enjoy superior candidacy for this status. This is Kim's 'causal exclusion argument', which is another example of an argument for a particular metaphysics of mind—this time a physicalist identity theory that identifies mental *properties* with physical *properties*—which focuses on the problem of mental causation.

Kim's causal exclusion argument has also been directed against 'non-reductive' physicalist views. According to non-reductive physicalism, that which is mental is not *identical* with anything physical; nevertheless, physical states, events and properties realise, constitute, compose or exhaustively determine mental states, properties and events. In other words, a one-way necessitation relation obtains between physical and mental states, events and properties. Kim (2005) argues that, if mental entities are both distinct from physical entities and causally responsible for some physical effects, then these physical effects must be overdetermined or the principle of causal closure must be false.[6] Non-reductive physicalists have responded by questioning Kim's key assumption that an effect cannot have two sufficient causes unless there is overdetermination. According

[6] See also Crane (1995) and Heil (2013).

to these responses, because of the one-way necessitation relation between mental and physical entities, it is wrong to see them as overdetermining causes of the same effect (Árnadóttir and Crane 2013; Bennett 2003; List and Menzies 2009; Shoemaker 2013). These responses grant that, if physical states, events and properties did not realise, constitute, compose or exhaustively determine mental states, properties and events then there would be overdetermination. Therefore, these responses assume that some form of physicalism is still required to reconcile the phenomenon of mental causation with principles about what actual causation is like, in this case the principle of causal closure and the assumption that causal overdetermination is not prolific. In this way, considerations about mental causation are still taken to support some kind of physicalism.

1.2 The relational understanding of mental causation

The arguments for physicalism outlined above are strong. It seems like the only way to deny their conclusion is to reject causal closure, accept prolific overdetermination or deny that there is mental causation. Few are willing to reject causal closure as this principle is thought to enjoy empirical support and prolific overdetermination is widely regarded as implausible. This seems to leave epiphenomenalism—the view that mental events are caused by physical events, for example events occurring in the brain, but mental events have no physical effects—as the only option left for the non-physicalist. However, before we declare physicalism the only acceptable metaphysics of mind, I think it is important to question whether the conception of mental causation assumed by arguments for physicalism is the right way to think about the place of mentality in the causal world.

In most discussions of the problem of mental causation, mental causation is presented as a causal relation between mental and physical *events*. Recall Davidson's principle of causal interaction: 'at least some mental events interact causally with physical events' (1970/2001: 208), or the following formulations of the first premise of the causal argument: 'all mental occurrences have physical effects' (Papineau 2001: 9); 'we think of mental and physical events as causally related' (Hopkins 1978: 223); and '[m]ental events have physical effects' (Noordhof 1999: 367). Sometimes, mental causation is presented as a causal relation that can hold between *states*. For example, in Anthony Skillen's (1984: 514) version of the causal argument for physicalism, he claims that '[o]f some physical events and states, mental events and states are causes'. Often, events, states and processes are thought of as being very similar in nature. For example, when David Armstrong proposes that mental states are states that are 'apt for bringing about a certain sort of behaviour', he notes that his use of the word 'state' is 'not meant to rule out "process" or "event"' (1968: 82). In most discussions of mental causation, events, states and processes are

thought of as three subclasses of the same general ontological category, and any differences between them do not affect their suitability to be the relata of a causal relation. I will call members of this general ontological category *items*. In most discussions of the problem of mental causation, mental causation is presented as a cause–effect relation where at least one of the relata is a mental item. I call this understanding of mental causation the *relational understanding of mental causation*.

> **Relational understanding of mental causation:** mental causation is mental items (events, processes or states) standing in causal relations to physical items (e.g. movements of a person's body).

What is important to notice about the relational understanding is that, as Jennifer Hornsby puts it, it presents mental causation as something 'we are supposed to think of as causation *by* the mental' (2015: 129). Or, as Tim Crane puts it, 'the arguments for physicalism must assume that the labels "mental" and "physical" as applied to causation are really transferred epithets—what is mental and physical are the relata of causation, not the causation itself' (1995: 219). Most discussions of mental causation thus assume that what is distinctive about mental causation is that it involves a mental relatum. For example, Thomas Kroedel writes that mental causation is 'the causation of physical effects by mental causes … there is mental causation whenever what is going on in our minds causes our bodies to move' (2020: 1). Philosophers writing about the problem of mental causation are limited to this way of describing what mental causation is because they assume that 'cause' is an unequivocal term—all causation everywhere is the same kind of thing, so the only thing that can discriminate between different categories of causation is the nature of the relata involved. This is the understanding of mental causation that has become standard in philosophy of mind. However, I believe this understanding of mental causation is misconceived.

The relational understanding of mental causation encourages us to accept an ontology of mental *items*. It presupposes that mental concepts pick out items, which can be causes and whose intrinsic nature is up for discovery. So, when someone believes something, an item called a belief exists. When someone wants something, an item called a desire exists. However, it is not obvious to me that the best way to describe mentality is as a collection or succession of mental items. I am sceptical that our status as minded creatures depends on the existence of mental items whose nature we have yet to discover and whose existence must, one way or another, be reconciled with the idea that the world is physical in all its fundamental aspects. On this matter, I agree with Gilbert Ryle. Ryle (1949) objects to what he calls the 'para-mechanical' view of the mind. According to this view, to have a mind—to have beliefs, desires, intentions, ideas, ambitions, emotions etc.—is to possess an inner causal mechanism that operates invisibly but which has observable actions as output. Mental states and

processes are conceived of as hidden inner causes of observable behaviour, and this, for Ryle, is a mistake. Ryle argues that when we use mental predicates we are not denoting (or even connoting) episodes in a person's secret history or alterations to an inner stream of consciousness. Furthermore, when we say that someone acted because of what she believed, knew, intended, desired, imagined, remembered or felt, these concepts do not—in fact, *could* not—refer to items that stand in causal relations to physical events. The idea that there can be mental causes is a category mistake, according to Ryle. Mental states and processes are just not the right sort of thing to be causes and effects.

I agree that mental states and processes are not the right sort of thing to be causes and effects. At the very least, it is wrong to assume that they can be causes in anything like the same sense in which the strike of a match can be the cause of its lighting, or the dropping of a glass can be the cause of its breaking. One reason for this is because it is a mistake to gloss over the differences between events, states and processes and assume that all three kinds of entity can be causes in the same way.

One assumption that is often made is that states and processes—or at least token states and token processes—are *particulars* just like events. However, this assumption is questionable. Particulars are *concrete, unrepeatable* entities. Helen Steward (1997) further suggests that something is a particular if it is capable of having a 'secret life' where an entity has a secret life if and only if:

1. the entity might be uniquely identified by means of some referring expression which is not known to apply to it by someone who is, nevertheless, in a position to single that entity out in some other way;
2. for some such referring expressions, the subject's not knowing that they provide an alternative means of uniquely identifying the entity in question is not simply a matter of her being ignorant of an alternative means of uniquely identifying some other entity;
3. for some such referring expression, the subject not knowing that they provide an alternative means of uniquely identifying the entity in question is not simply a matter of her not knowing about one of the entity's relational properties (where spatial and temporal properties are not accounted relational). (1997: 32)

Entities that satisfy these conditions are entities that can be uniquely referred to with more than one expression, each of which picks out the entity via a different intrinsic feature of it. For example, the morning star and the evening star are one and the same entity, the planet Venus, but each expression picks out Venus via different intrinsic features of it: appearing in the sky in the morning and appearing in the sky in the evening. It is because each expression refers to Venus via different intrinsic features that identifying the morning star with the evening star is informative and not trivially true. Venus satisfies the secret life requirement; therefore, it is a particular.

Steward argues that entities that have a propositional structure, like facts, cannot be particulars because they do not meet the secret life requirement. Steward (1997) argues convincingly that states also have a propositional structure and so are also not particulars. I also do not think that processes are particulars. I think they are universals—they are single repeatable entities. I will have more to say about this in Chapter 6 (see also White 2020). This means that, even if mental states and processes are cited in causal explanations, we cannot assume that the role they play in such explanations is exactly the same as the role we would take an event to play.

Denying the existence of any mental particulars at all is too strong to be plausible. Suddenly remembering something, successfully imagining something, noticing something, realising something all seem to be mental events and all seem to exist. However, in Rylean spirit, I think these events should not be thought of as occurring in a 'secret history'. Such events are often thought of as episodes of a mental succession—and this is what I believe is wrong. Although these events are mental, they are, in a very banal sense, physical events too because they are all actions of human beings. These events should be dealt with as actions—not as happenings in a private mental sequence, whose intrinsic nature is unknown to us.

I do not want to say there is no such thing as mental causation. I am sceptical of the idea that there are mental *items* that stand in causal relations but I do think there is causation that deserves to be called 'mental'. I believe that the causal processes human beings engage in when they act intentionally count as mental causation, and the mentality of these causal processes consists in the fact that these processes are part of a larger pattern of meaningful, or interpretable, activity. One aim of the later chapters of this book is to prove this.

An example might help to summarise why I am sceptical of the relational understanding of mental causation. Some years ago, my partner was driving along a country road at night when, from the passenger's seat, I spotted a deer in the road in front of us. I immediately called out "deer, deer, deer!" to alert the driver. I have three intuitions about this example. First, the idea that the event-causal sequence from first noticing the deer to calling out involves a mental-but-not-physical event as a causal intermediary seems incorrect. My intuition is that, if we were to describe the example in event-causal terms we would mention light entering my eye, electrical activity in my brain, contractions of my muscles—all events that we would class as physical—and that's all. Second, the idea that one of the physical events in the chain just described, or even a subset of those physical events, is identical with, or somehow constitutes the mentality of the example also seems wrong. It seems wrong to suppose that any of the physical events in the causal sequence amounts to my conscious experience, or to my worry that we might hit a deer, or to my desire that the driver be alerted, or any of the mental experiences I think are present in the example. (In other words, I do not think physicalism is true.) Third, I think there *is* mental causation in this case—I did as I did because of what I thought and experienced.

The puzzle is: how can all these intuitions be true? The solution, I suggest, is to rethink mental causation. What makes this case an example of mental causation is not that there is a causally efficacious mental item that caused my behaviour.

1.3 Mental causation and human agency

I believe that the relational understanding of mental causation is presupposed in many debates within philosophy of mind because of a triad of popular, and individually plausible, philosophical theories: physicalism, causal theories of intentional action and a relational approach to causation. I call this triad of views *the physicalist triad*. Although these theories are logically independent and about distinct philosophical questions, in practice they are mutually reinforcing.

To see this, we should first ask why the relational understanding of causation has become the standard way of thinking about mental causation within philosophy of mind. As I have already mentioned, few philosophers writing on the problem of mental causation are willing to deny the existence of mental causation. However, there is an important exception to this norm. 'Epiphenomenalism' is the view that mental items are caused by physical items, for example events occurring in the brain, but mental items have no physical effects. Epiphenomenalism can be seen as a response to the causal argument for physicalism. The argumentative force of the causal argument is that if mental and physical items are distinct, then they are in competition with each other for status as the cause of a physical effect, for example a bodily movement. The first premise of the causal argument states that mental items have such physical effects; the second premise says that every physical effect has a sufficient physical cause, so either the physical effects like bodily movements are overdetermined, or the mental and physical causes are one and the same, or we will have to accept that one of the candidate causes is not really a cause after all. It is often assumed that, if you are pushed towards the last option, you have to admit that it is the mental candidate that turns out not to be a cause. This is precisely what epiphenomenalists do.

Epiphenomenalism is a minority position, however. Most philosophers of mind take it as a position to be avoided. It will be useful to briefly consider why this is, as it provides evidence to think that the relational understanding of causation has become the dominant understanding of mental causation because of intuitions about human agency.

First, consider Kim's remarks on why it is important that mental causation is real:

> First and foremost, the possibility of human agency, and hence our moral practice, evidently requires that our mental states have causal effects in the physical world. In voluntary actions our beliefs and

> desires, or intentions and decisions, must somehow cause our limbs to move in appropriate ways, thereby causing the objects around us to be rearranged. (2005: 9)

Here Kim endorses the idea that the possibility of human agency depends on beliefs and desires, or intentions and decisions, standing in causal relations to bodily movements. I agree that some form of mental causation is indispensable to our conception of ourselves as agents who act intentionally and bear moral responsibility. I believe that such a conception presupposes the reality of 'causal processes involving mental phenomena', as Menzies (2013: 58) puts it. However, these claims are much weaker than the claim Kim makes. Kim claims that the possibility of human agency depends on beliefs, desires, intentions and decisions somehow causing our limbs to move.

Kim also states that:

> [I]t seems plain that the possibility of psychology as a science capable of generating law-based explanations of human behaviour depends on the reality of mental causation: mental phenomena must be capable of functioning as indispensable links in causal chains leading to physical behaviour, like movements of the limbs and vibrations of the vocal cord. A science that invokes mental phenomena in its explanations is presumptively committed to their causal efficacy; if a phenomenon is to have an explanatory role, its presence or absence must make a difference—a causal difference. (2005: 10)

Again, I agree that the worth of psychology as a science and as a means by which we can predict, explain and control each other's behaviour requires that people's behaviour can be causally explained by what they think, feel, believe and want. However, again, this is a weaker claim than Kim's. Kim claims that the possibility of psychological explanations of human behaviour requires that 'mental phenomena must be capable of functioning as indispensable links in causal chains leading to physical behaviour, like movements of the limbs and vibrations of the vocal cord'. Thus, Kim thinks the possibility of psychological explanation presupposes that mental phenomena, like believing that one ought to brush one's teeth or wanting to make a cup of tea, are *links in causal chains*.

Kim's justification for believing in mental causation seems to be that, unless mental items stand in causal relations to physical items, human agency and psychological explanation of human behaviour would be impossible. Kim also considers perceptual knowledge as evidence that mental causation—this time causation *of* mental events, as opposed to causation *by* mental events—must exist. The thought here is that if the content of perceptual experience is to indicate what the world is like then perceptual experiences must be caused by external states of affairs. However, this kind of mental causation is rarely the focus in debates about physicalism. Furthermore, this kind of mental

causation could not serve as a premise in the causal argument for physicalism described above. This is because the causal closure principle concerns the causes of physical *effects*, so the mental causation it potentially excludes is causation *of* physical effects *by* mental items.

It is not just Kim who thinks that human agency is possible only if mental items stand in causal relations to physical events such as bodily movements. The reaction to the work of Benjamin Libet shows that many others share Kim's belief. Libet (1985) conducted an experiment where participants were asked to move their finger when they felt like it and note the time at which they felt an urge to move their finger. While participants did this, Libet recorded their brain activity and found that a particular signal, known as a readiness potential, was correlated with the participants' finger movements and, significantly, occurred 350 milliseconds before participants reported having an urge to move their finger. This implied that the urge to move their finger could not have initiated the finger movement, and therefore that 'conscious will' does not play the role in 'voluntary action' that we think it does (Libet 1999; see also Libet 2002).

Libet thus assumes that for voluntary action to be possible, the action must be initiated by a mental item—in this case a conscious act of will. Similar work by John-Dylan Haynes and Michael Pauen (2013) found that neural activity correlated with the outcome of a choice whether to add or subtract digits occurred before participants recorded consciously making the decision. Haynes and Pauen take this to show that even more abstract choices, such as whether to do an addition or a subtraction, are not initiated by conscious mental events. Since Libet's early experiments, many more experiments on neurological preparatory processes for movement have been conducted, with neuroscientists asking what 'action-related cognitive processes' might be 'encoded' by such neurological activity (Fifel 2018: 785).

Libet's experiment, and those like it, have been criticised on the grounds that the task participants are being asked to do is artificial: participants are instructed to decide to do something spontaneously, hence extrapolation to other kinds of voluntary action, which might involve much more complicated or extended deliberation, is not justified. Other philosophers have argued that these kinds of experimental findings pose no threat to the possibility of voluntary action for which we can be held responsible, because there are still elements of the action performed in Libet's (and presumably Haynes's) experiment that are initiated by conscious mental events. For example, Owen Flanagan (1996) argues that, as long as taking part in Libet's experiment was consciously initiated, it does not matter if the realisation of this 'big picture' decision was not consciously initiated. It has also been suggested that as long as the precise details of the movement (e.g. whether to use one's left or right hand) are consciously initiated then it does not matter if the action as a whole was unconsciously initiated (Haggard and Libet 2001). However, these responses implicitly accept that an action needs to be initiated by a conscious mental event for it to count as free or voluntary or the kind of action that the agent would be responsible for.

Neil Levy (2005) challenges this assumption, arguing that the idea that only consciously initiated actions can be free is conceptually incoherent. Reaching a decision, Levy argues, cannot be something we consciously control. Deliberation may be a conscious activity but decisions themselves are not 'actions performed by consciousness', Levy claims; they are, rather, events we wait for and passively witness. Levy argues that decisions cannot be events we consciously control for the following reason:

> [D]ecision making is, or is an important element of, our control system, whereby we control our activity and thereby attempt to control our surroundings. If we were able to control our control system, we should require another, higher-order, control system whereby to exert that control. And if we had such a higher-order control system, the same problems would simply arise with regard to it. The demand that we exercise conscious will seems to be the demand that we control our controlling. And that demand cannot be fulfilled. (2005: 73)

I have a lot of sympathy with Levy's argument that there is something conceptually confused about the idea that decisions are conscious mental events which initiate voluntary actions. This is because I do not think it is obvious that the possibility of free or voluntary action depends on mental events being initiators of bodily movements. Those tempted to see Libet-style experiments as threatening to our conception of ourselves as capable of free action seem to presuppose free or voluntary action is only possible if mental items stand in causal relations to bodily movements.

Sophie Gibb also seems to endorse the idea that our conception of human agency entails that beliefs and desires stand in causal relations to bodily movements. She writes that:

> The thought that mental causes have physical effects—that our beliefs and desires can give rise to the movement of our bodies—is central to our pre-theoretical notion of human agency. (2013: 321)

What is interesting about Gibb's remark is that she also claims that this idea is a 'pre-theoretical notion'. This is a claim that I find surprising. Our everyday pre-theoretical way of talking about the mind and its place in the causal world does not usually involve talking about mental causes. We talk often about doing things because of what we believe, think, want or feel, but most of the time these discussions do not obviously involve identifying something that occurred at a particular time that triggered our action, or which moved us from a state of inaction to a state of action. As Elizabeth Anscombe noted, when 'one says what desire an act was meant to satisfy, one does not identify a feeling, image or idea that precedes the act the desire explains: one does not answer the question "what did you see or hear or feel, or what ideas or images cropped up in your

mind and led up to it?'" (2000: 17). This is not to say that we *never* speak about mental causes—of course we do. My claim is rather that a lot of talk about how our thoughts and feelings feature in our daily lives does not mention mental causes. There may be reasons to think that when we explain our actions in terms of our thoughts and feelings such explanations can only be true if mental causes have physical effects (I will examine those reasons in Chapter 2), but this would constitute a philosophical argument not a pre-theoretical notion. Indeed, it is precisely this philosophical argument that I believe is the reason the relational understanding of mental causation is so widely endorsed.

Within the philosophy of mind, and especially in discussions concerning physicalism, many believe that intentional or voluntary human action is possible only if mental items stand in causal relations to physical events such as bodily movements. The idea that psychological explanations of intentional action, and even the possibility of intentional action itself, entails the existence of causal relations between mental items and physical items is central to a view of intentional action that I call the 'causal theory of intentional action'. This theory of intentional action has become the dominant theory of intentional action. Although this theory concerns what makes an action intentional, and is logically independent from physicalism, I believe that implicit commitment to this theory is what makes the relational understanding of mental causation seem undeniable to physicalists. Physicalists, and those sympathetic to physicalism, assume the relational understanding of mental causation because of implicit acceptance of causal theories of intentional action. This is the first way in which different elements of the physicalist triad support each other.

1.4 Mental causation and the relational approach to causation

Physicalism and the relational approach to causation are also mutually reinforcing. This is because both views are thought to be *naturalistic*. Physicalism is often hailed as a naturalistic account of the mind, meaning that it is a metaphysics of mind that fits comfortably with a scientific view of the world, and especially a scientific view of causation. The thought is that physicalism does not commit us to the existence of anything that would be regarded as an unnatural addition to the world as described by science. Rightly or wrongly, the relational approach to causation is also part of this naturalistic worldview.

The connection between naturalism and the relational approach to causation is, I think, part of David Hume's influence on the philosophy of causation. Inspired by Hume's ideas about causation, philosophers of causation have often assumed that empirical science cannot provide us with any knowledge of necessitating connections in nature, whereby an object with certain powers 'must' behave in certain ways in certain conditions. Powers have long been regarded as epistemically suspicious and ineffable: we can perceive a thing's properties, what it is like, but not what it is capable of doing. Many philosophers of

causation therefore believe that causation, as it exists in reality, cannot be the exercise of power or efficacy or the making-happen of events if empirical science is to provide us with any knowledge of it. Hume himself tried to articulate a conception of causation divested of any association with natural necessity and did so by describing causation as a relation. Since then, philosophy of causation has proceeded under the assumption that there is one sort of thing that is causation: causation is a type of relation that holds independently of how the relata are described. To entertain the possibility that causation might be something richer than this, or that causation might refer to different things depending on the explanatory context, is to construe causation as something mysterious, ineffable and empirically unrespectable.

The mutually supportive connection between physicalism and the relational approach to causation can also be seen if we examine the principle of causal closure more closely. The principle of causal closure is a key part of any naturalistic account of the mind. If your account of the mind appeals to causal relations, and you want your account to be naturalistic, then the principle of causal closure is a constraint on what those causal relations can be like. The principle of causal closure is a key part of naturalism because it is supposed to be derived from, and supported by, the findings of scientific investigations into causal processes. Scientific discoveries are supposed to show that going beyond the physical realm to causally account for physical effects is unnecessary. The fact that physical science has been able to posit physical causes of many physical effects, and has never needed to 'leave the realm of the physical to find a fully sufficient cause' (Papineau 2001: 8), constitutes inductive evidence that every physical effect has a physical cause.

However, despite being readily assented to, the exact content of the causal closure principle is not obvious, as it is not clear what is meant by the terms 'physical' and 'sufficient cause'. Does it, for instance, mean that no physical effect can have as its cause a 'supernatural' substance? Or that no physical effect can have as its cause an event of a type that is not part of the subject matter of a physical science? These two interpretations are very different.

The first takes the referent of 'sufficient cause' to be a substance and 'physical' to mean anything that is not 'supernatural'. This version of the causal closure principle speaks against Cartesian minds being the causes of physical effects, it also seems to be empirically supported, but it does not immediately rule out that there might be physical effects that have non-physical *events* as causes. As Barry Stroud points out, the existential claim that the world contains only non-supernatural substances allows the character of the world to be 'as rich as you please' (1986: 264). Including irreducible, causally efficacious mental states and events is not at all problematic as long as they are states of, or events involving, non-supernatural beings.

The second interpretation takes the referent of 'sufficient cause' to be an event and 'physical' to mean strictly expressible in the vocabulary of physical science, which is a stronger thesis and more useful to the physicalist argument.

However, it is more difficult to justify empirically. A valid generalisation from the causal processes science investigates to all causal processes depends on the causal processes science investigates being suitably similar to other causal processes, including those that are supposed to involve mental causation. As we saw in the previous section, most arguments for physicalism focus on the mental causation associated with human action. However, are human actions suitably similar to the causal processes physical science investigates? Human actions are sometimes directed towards a goal, they can succeed or fail, they can be paused and then continued, and they can be started but never completed. In many ways, human actions are unlike the causal processes investigated by physical science; therefore, it is at least questionable whether a principle like causal closure should apply to them. However, this line of thought is unlikely to be persuasive if you are sympathetic to a relational approach to causation. According to the relational approach to causation, all causation is the same; causation is a homogenous phenomenon. Causation can hold between diverse relata, but there is one relation of causation, and that is all causation is. On this approach, then, the features I just ascribed to human actions should be analysable in terms of an event-causal sequence. They represent ways in which the causality of human action might involve unique relata (neural events maybe)—they do not represent unique features of the causation itself. Therefore, features such as goal-directedness do not represent properties that radically distinguish human actions from the kinds of causal processes investigated by physical science. All causation is the same; only the relata change. In this way, the relational approach to causation supports physicalism.

1.5 Troubles with physicalism

Physicalism has dominated philosophy of mind. It can seem like the only plausible option when faced with arguments centred around mental causation. However, I believe the reason for this dominance is because of physicalism's connection to causal theories of intentional action and relational approaches to causation and not because physicalism itself is the best, most informative metaphysics of the mind. As I said in the introduction, it is not my intention to argue directly against physicalism in this book. I do not intend to prove that physicalism must be false. This is because my main objection to the physicalist triad is not that physicalism itself is implausible or internally inconsistent. What I object to is physicalism's *hegemony*. Physicalism often seems like the only option. Debates can be had about what kind of physicalist you want to be, but the causal argument seems to make it impossible to doubt that some form of physicalism must be true. This is where my discomfort lies, as I believe the arguments for physicalism, the arguments that make physicalism seem like such an indubitable theory, are artificially bolstered by their association with causal theories of intentional action and relational approaches to causation.

Physicalism can only be given a fair assessment after the conception of mental causation assumed by arguments for physicalism—the conception of mental causation suggested by causal theories of intentional action and relational approaches to causation—is examined. Nevertheless, it is useful to say something about why someone might be dissatisfied with physicalism as a metaphysics of mind, as this will help justify my project.

First, there are those who doubt the coherence of physicalism. Physicalism says everything that exists is either itself a physical entity or 'nothing over and above' a physical entity—but what exactly is a physical entity? For some, this interpretive difficulty represents a fundamental flaw with physicalism as a metaphysical theory. It shows either that physicalism must be false or that it does not make a substantive claim.

In everyday discourse, we might take physical to mean 'relating to things we can sense' or 'extended in space' or 'not supernatural'. This everyday conception, however, is not precise enough to be useful in a formulation of physicalism. To say that everything that exists is related to things we can sense is very vague—it could potentially include mental states, properties and events, as these are related to human beings, and those are the kind of entity the physicalist wants to say do not fundamentally exist. To say that everything that exists is extended in space makes physicalism indistinguishable from materialism, the view that everything that exists is matter or made of matter. Materialism is no longer a popular view among philosophers as modern physics seems to countenance the existence of entities that are not matter or made of matter—photons, for example. To say that everything that exists is not supernatural rules out the existence of souls or purely mental substances but leaves room for the existence of irreducibly mental states, properties and events as long as they are states and properties of, or events involving, non-supernatural beings.

For many physicalists, specifying what it is to be physical must include reference to physical science. One way to do this is to define the physical entities as those which physical sciences tell us about. However, now we face the problem of deciding which sciences count as physical sciences. Physics obviously counts. Does biology? Does economics? Does psychology? There must be some way of drawing a boundary between physical sciences and non-physical sciences, otherwise the claim that everything is physical will be trivially true. One possibility is that physical sciences are those whose laws could be expressed in the vocabulary of *physics* with the help of suitable translation principles. Some sciences could not be reduced in this way—those sciences are the non-physical sciences. However, there is a famous issue with this suggestion. Carl Hempel (1969) pointed out that, if 'physical' is defined in terms of contemporary physics, then physicalism must be false as we cannot presume that contemporary physics is complete and describes everything that exists. However, if 'physical' is defined in terms of a future, complete physics, then physicalism is trivially true, because for all we know a future physics might include entities that the physicalist now wants to class as non-physical, namely mental entities. This is Hempel's dilemma.

In response to Hempel's dilemma, physicalists have suggested several other ways to define what 'physical' means. One way is to define physical in terms of what it is not: 'Physical entities are not fundamentally mental (that is, do not individually possess or bestow mentality)' (Wilson 2006: 61). This method of defining the physical allows physicalists to continue to use physics as the authority when it comes to describing the fundamental ontology of the world, acknowledging that current physics is incomplete and therefore that the world may contain more than current physics says it contains. However, the method avoids making physicalism trivially true because, although we do not know exactly what the ontology of a complete physics will be, we can say that it won't include fundamentally mental entities because fundamentally mental entities are not physical. Physicalism is a theory about what fundamentally exists and, as Papineau puts it, to do that 'it isn't crucial that you know exactly what a complete physics would include. Much more important is to know what it won't include' (2001: 12).

Another way to define 'physical' is inspired by structural realism, the view that we should believe to be true the structural content of our best scientific theories. According to this view, physical entities are those which are described by the mathematical language characteristic of physics. This means that the as-yet-undiscovered entities of future physics will count as physical because they will also be describable in this mathematical language (thus Hempel's dilemma is avoided) (Chalmers 2020).

Alyssa Ney (2008) further suggests that the best way to understand the content of physicalism is not to understand it as a truth-apt statement about what exists (and so not as the kind of statement that could be subject to Hempel's dilemma) but to see it as an 'oath' to formulate one's ontology according to physics. At least for now, physics does not include fundamentally mental entities; therefore, physicalists do not believe in fundamentally mental entities.

This is not an exhaustive list of all the ways physicalists have tried to solve the problem of specifying what is 'physical'. Suffice it to say that the debate is ongoing. For some, this issue is unsolvable. Tim Crane and D. H. Mellor argue that 'physicalism lacks a clear and credible definition, and that in no non-vacuous interpretation is it true' (1990: 394). A key premise of their argument is that, for physicalism to stand a chance of being non-vacuously true, there must be a reason why physical science is a 'unique ontological authority' (1990: 394). In other words, there must be a credible answer to the question 'what is special about physical science such that it can tell us what exists, whereas non-physical sciences cannot?' Crane and Mellor are of the opinion that there is no credible answer to this question.

I will not give a detailed assessment of Crane and Mellor's argument, but there are two interesting points they make in their 1990 paper that I want to highlight. The first is that physics might seem like an ontological authority because it can seem *universal*, i.e. about everything, and *basic*, i.e. about the most fundamental parts of the world. Crane and Mellor suggest that physics seems universal and basic because of its conventional subject matter. For example,

physics encompasses the study of the fundamental particles that compose all paradigmatically physical objects from rocks to animals to planets. Physics also encompasses mechanics and so studies everything that moves. A science with such vast reach can seem *universal*, and a science that studies the microscopic building blocks of the world can seem *basic*. However, just because a science has a far-reaching subject matter does not mean everything falls within its domain and just because a science studies the microscopic building blocks of a great many things does not mean it studies everything that exists.

The second point of interest is Crane and Mellor's argument against the suggestion that physical sciences enjoy ontological authority because only physical sciences describe causal reality. Crane and Mellor's view here is particularly interesting given the connections I believe exist between physicalism and certain assumptions about the nature of causation. Crane and Mellor argue that there is no good reason to deny that non-physical sciences like psychology describe causal reality. One bad reason (in their opinion) is the stipulation that for a causal claim to be true it must depend on an extensional causal relation between particulars. Crane and Mellor argue that there are examples within physics that show that this stipulation is false, and so there is no reason to deny that true causal claims can be made within psychology even though these claims often do not rest on an extensional causal relation between particulars. This discussion is interesting because it is another example of how physicalism derives support from substantive assumptions about the nature of causation. When those assumptions are questioned, the case for being a physicalist is weakened. It is also interesting because in Chapter 7 I will make a similar argument about the nature of causal explanations. I will argue that it is not necessary for an explanation to be causal that its explanandum designate an effect and its explanans designate an item which is the cause of that effect. In so doing, I will allow causal explanations that use mental concepts to count as revealing causal truths even though they do not designate particulars which are causes and particulars which are effects.

There are many options open to the physicalist when it comes to specifying what 'physical' means. It would be very presumptive of me to say none of those options is satisfactory; I cannot think of any *a priori* reason why one of the many options cannot be made to work. However, I think this interpretive difficulty shows that physicalism's credentials as an informative metaphysical theory are not as secure as they may first appear. A lot depends on how physicalism's core thesis is made precise.

A second major criticism of physicalism is that there are questions about the mind that physicalism cannot answer and for this reason it is an unsatisfactory account of mentality. The most famous version of this critique is David Chalmers's (1996) so-called 'hard problem of consciousness'. Chalmers's argumentative strategy is as follows. There are strong reasons for thinking that it is impossible to explain consciousness in physical terms. The idea that one day a scientist could

look at a person (or maybe their brain) and say, "they are having this conscious experience because they have these physical properties" and that constitute a fully satisfactory explanation is not credible (at least, according to Chalmers). Even if we imagine that every physical fact about a person is fully specified, it seems like we could still wonder why that person is having the conscious experience they are having. There is an 'explanatory gap' (Levine 1983). From this explanatory gap, Chalmers concludes that there is also a metaphysical gap. The reason we cannot explain how consciousness arises from physical states is because it does not arise from physical states. Physicalism can never be true of mental states that have a phenomenal quality, i.e. states that we consciously experience.

A vast literature exists on this topic. There are those who argue that drawing a metaphysical conclusion from an explanatory problem is unjustified (Hill 1997; Yablo 1993), there are those who argue that the explanatory gap problem is misconceived (Tye 1999) or that it is not as insurmountable as it seems (Dennett 2005), and there are those who argue that an explanatory gap is to be expected if physicalism is true and so actually confirms rather than disproves physicalism (Papineau 2002). I do not intend to get into this debate here. I bring up the issue of the hard problem to make a simpler point. The mind–body problem —the central issue in philosophy of mind—concerns the question of how physical things like human beings are capable of thought (including conscious thought). The explanatory gap seems to indicate that the mind–body problem is not satisfactorily answered by physicalism. Physicalism sidesteps the mind–body problem rather than directly confronting the issue of how our physical construction enables us to engage in mental activities like thinking, imagining and feeling. The fact that brain activity seems to have some connection to our ability to think is a genuinely puzzling fact—brain activity and thinking seem totally unlike each other. Physicalism seems to ignore rather than engage with that puzzlement. For this reason, there is an incentive to at least consider other options. So, even if Chalmers's argument against physicalism does not actually prove that physicalism is false, the hard problem gives us reason to consider alternative metaphysical accounts of the mind.

I have outlined some criticisms of physicalism discussed in philosophy of mind. I have not adjudicated on any of these issues. I think the problems outlined above are serious but it would take a much more detailed analysis to assess whether the problems are insurmountable and therefore falsify physicalism. Nevertheless, the foregoing discussion shows that physicalism is not without its critics. It is not a perfect metaphysics of mind; therefore, the fact that it can often seem like the only option is a problem. I believe that the physicalist triad has limited our thinking about mental causation and therefore prevented us from exploring more diverse accounts of the relationship between our mind and body. The following chapters investigate whether there is a way to break out of this triad, and thereby open up new ways of understanding mental causation. It is my hope that doing this will refresh debates in philosophy of mind

by allowing us to postulate new metaphysical theories of mentality that may be able to answer questions about the mind that physicalism cannot.

References[7]

Anscombe, G E M 2000 *Intention*. Cambridge, MA: Harvard University Press. Originally published in England in 1957 by Basil Blackwell.

Armstrong, D M 1968 *A materialist theory of the mind*. London: Routledge.

Árnadóttir, S T and Crane, T 2013 There is no exclusion problem. In: Gibb, S C, Lowe, E J and Ingthorsson, R *Mental causation and ontology*. Oxford: Oxford University Press. pp. 248–266.

Bennett, K 2003 Why the exclusion problem seems intractable and how, just maybe, to tract it. *Noûs*, 37(3): 471–497. DOI: https://doi.org/10.1111/1468-0068.00447

Bryant, A 2020 Physicalism. In: Raven, M J *The Routledge handbook of metaphysical grounding*. New York: Routledge. pp. 484–500.

Burge, T 1979 Individualism and the mental. *Midwest Studies in Philosophy*, 4(1): 73–122. DOI: https://doi.org/10.1111/j.1475-4975.1979.tb00374.x

Chalmers, D 1996 *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.

Chalmers, D 2020 Is the hard problem of consciousness universal? *Journal of Consciousness Studies*, 27(5–6): 227–257.

Crane, T 1995 The mental causation debate. *Aristotelian Society Supplementary Volume*, 69(1): 211–236. DOI: https://doi.org/10.1093/aristoteliansupp/69.1.211

Crane, T and Mellor, D H 1990 There is no question of physicalism. *Mind*, 99(394): 185–206. DOI: https://doi.org/10.1093/mind/xcix.394.185

Davidson, D 1970 Mental events. In: Foster, L and Swanson, J W *Experience and theory*. 2nd ed. Cambridge, MA: University of Massachusetts Press. Reprinted in Davidson 2001 pp. 207–224.

Davidson, D 2001 *Essays on actions and events*. 2nd ed. Oxford: Clarendon Press.

Dennett, D 2005 *Sweet dreams: Philosophical obstacles to a science of consciousness*. Cambridge, MA: MIT Press.

Elpidorou, A and Dove, G 2018 *Consciousness and physicalism: A defense of a research program*. New York: Routledge.

Fifel, K 2018 Readiness potential and neuronal determinism: New insights on Libet experiment. *Journal of Neuroscience*, 38(4): 784–786. DOI: https://doi.org/10.1523/JNEUROSCI.3136-17.2017

[7] Author note: some references to Davidson are formatted (1963/2001). This indicates the initial date of publication of the paper (in this case 1963) but references the paper as it appears in the 2001 collection of his essays, with the page numbers relating to that volume.

Flanagan, O J 1996 *Self expressions: Mind, morals, and the meaning of life*. Oxford: Oxford University Press.

Gibb, S 2013 Introduction to mental causation and ontology. In: Gibb, S C, Lowe, E J and Ingthorsson, R *Mental causation and ontology*. Oxford: Oxford University Press. pp. 1–17.

Haggard, P and Libet, B 2001 Conscious intention and brain activity. *Journal of Consciousness Studies*, 8(11): 47–63.

Haugeland, J 1982 Weak supervenience. *American Philosophical Quarterly*, 19(1): 93–103. http://www.jstor.org/stable/20013945

Haynes, J-D and Pauen, M 2013 The complex network of intentions. In Caruso, G D *Exploring the illusion of free will and moral responsibility*. Plymouth: Lexington Books. pp. 221–238.

Heil, J 2013 Mental causation. In: Gibb, S C, Lowe, E J and Ingthorsson, R *Mental causation and ontology*. Oxford: Oxford University Press. pp. 18–35.

Hellman, G P and Thompson, F W 1975 Physicalism: Ontology, determination, and reduction. *Journal of Philosophy*, 72(17): 551–564. DOI: https://doi.org/10.2307/2025067

Hempel, C 1969 Reduction: Ontological and linguistic facets. In White, M et al. *Philosophy, science, and method: Essays in honor of Ernest Nagel*. New York: St Martin's Press. pp. 179–199.

Hill, C S 1997 Imaginability, conceivability, possibility and the mind-body problem. *Philosophical Studies*, 87(1): 61–85. DOI: https://doi.org/10.1023/a:1017911200883

Hopkins, J 1978 Mental states, natural kinds and psychophysical laws. *Aristotelian Society Supplementary Volume*, 52(1): 195–236. DOI: https://doi.org/10.1093/aristoteliansupp/52.1.195

Horgan, T E 1993 From supervenience to superdupervenience: Meeting the demands of a material world. *Mind*, 102(408): 555–586. DOI: https://doi.org/10.1093/mind/102.408.555

Hornsby, J 2015 Causality and 'the mental'. *HUMANA.MENTE Journal of Philosophical Studies*, 8(29): 125–140.

Jackson, F 1998 *From metaphysics to ethics: A defence of conceptual analysis*. Oxford: Oxford University Press.

Kim, J 1989 The myth of nonreductive materialism. *Proceedings and Addresses of the American Philosophical Association*, 63(3): 31–47. DOI: https://doi.org/10.2307/3130081

Kim, J 1993 *Supervenience and the mind: Selected essays*. Cambridge: Cambridge University Press.

Kim, J 1998 *Mind in a physical world*. Cambridge: Cambridge University Press.

Kim, J 2001 Mental causation and consciousness: The two mind-body problems for the physicalist. In: Carl, G and Loewer, B *Physicalism and its discontents*. Cambridge: Cambridge University Press. pp. 271–283.

Kim, J 2005 *Physicalism, or something near enough*. Princeton, NJ: Princeton University Press.

Kroedel, T 2020 *Mental causation: A counterfactual theory*. Cambridge: Cambridge University Press.

Levine, J 1983 Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64(October): 354–361. DOI: https://doi.org/10.1111/j.1468-0114.1983.tb00207.x

Levy, N 2005 Libet's impossible demand. *Journal of Consciousness Studies*, 12(12): 67–76.

Libet, B 1985 Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8(4): 529–566. DOI: https://doi.org/10.1017/s0140525x00044903

Libet, B 1999 Do we have free will? *Journal of Consciousness Studies*, 6(8–9): 47–57.

Libet, B 2002 The timing of mental events: Libet's experimental findings and their implications. *Consciousness and Cognition*, 11(2): 291–299. 10.1006/ccog.2002.0568

Lowe, E J 2000 Causal closure principles and emergentism. *Philosophy*, 75(294): 571–586. DOI: https://doi.org/10.1017/s003181910000067x

List, C and Menzies, P 2009 Nonreductive physicalism and the limits of the exclusion principle. *Journal of Philosophy*, 106(9): 475–502. DOI: https://doi.org/10.5840/jphil2009106936

Lynch, M P and Glasgow, J 2003 The impossibility of superdupervenience. *Philosophical Studies*, 113(3): 201–221. DOI: https://doi.org/10.1023/a:1024037729994

Melnyk, A 2003 *A physicalist manifesto: Thoroughly modern materialism*. New York: Cambridge University Press.

Melnyk, A 2006 Realization and the formulation of physicalism. *Philosophical Studies*, 131(1): 127–155. DOI: https://doi.org/10.1007/s11098-005-5986-y

Melnyk, A 2018 In defense of a realization formulation of physicalism. *Topoi*, 37(3): 483–493. DOI: https://doi.org/10.1007/s11245-016-9404-1

Menzies, P 2013 Mental causation in the physical world. In: Gibb, S. C., Lowe, E. J., and Ingthorsson, R. D. *Mental causation and ontology*. Oxford: Oxford University Press. pp. 58–88.

Morris, K 2018 Truthmaking and the mysteries of emergence. In Vintiadis, E and Mekios, C *Brute facts*. Oxford: Oxford University Press. pp. 113–129.

Ney, A 2008 Defining physicalism. *Philosophy Compass*, 3(5): 1033–1048. DOI: https://doi.org/10.1111/j.1747-9991.2008.00163.x

Noordhof, P 1999 The overdetermination argument versus the cause-and-essence principle—no contest. *Mind*, 8(430): 367–375. DOI: https://doi.org/10.1093/mind/108.430.367

Papineau, D 1993 *Philosophical naturalism*. Oxford: Blackwell.

Papineau, D 2001 The rise of physicalism. In: Gillett, C and Barry, L *Physicalism and its discontents*. Cambridge: Cambridge University Press. pp. 3–26.

Papineau, D 2002 *Thinking about consciousness*. New York: Oxford University Press.

Pettit, P 1993 A definition of physicalism. *Analysis*, 53(4): 213–223. DOI: https://doi.org/10.1093/analys/53.4.213

Putnam, H 1975 The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science*, 7: 131–193.

Ryle, G 1949 *The concept of mind*. London: Hutchinson's University Library.

Shoemaker, S 2001 Realization and mental causation. In Gillett, C and Loewer, *The proceedings of the Twentieth World Congress of Philosophy*. Cambridge: Cambridge University Press. pp. 23–33.

Shoemaker, S 2007 *Physical realization*. Oxford: Oxford University Press.

Shoemaker, S 2013 Physical realization without preemption. In: Gibb, S C, Lowe, E J and Ingthorsson, R *Mental causation and ontology*. Oxford: Oxford University Press. pp. 35–57.

Skillen, A 1984 Mind and matter: A problem which refuses dissolution. *Mind*, 93(372): 514–526. DOI: https://doi.org/10.1093/mind/xciii.372.514

Steward, H 1997 *The ontology of mind: Events, processes and states*. Oxford: Oxford University Press.

Stroud, B 1986 The physical world. *Proceedings of the Aristotelian Society*, 87(1): 263–277. DOI: https://doi.org/10.1093/aristotelian/87.1.263

Sturgeon, S 1998 Physicalism and overdetermination. *Mind*, 107(426): 411–432. DOI: https://doi.org/10.1093/mind/107.426.411

Tye, M 1999 Phenomenal consciousness: The explanatory gap as a cognitive illusion. *Mind*, 108(432): 705–725. DOI: https://doi.org/10.1093/mind/108.432.705

White, A 2020 Processes and the philosophy of action. *Philosophical Explorations*, 23(2): 112–129. DOI: https://doi.org/10.1080/13869795.2020.1753801

Wilson, J 2005 Supervenience-based formulations of physicalism. *Noûs*, 39(3): 426–459. DOI: https://doi.org/10.1111/j.0029-4624.2005.00508.x

Wilson, J 2006 On characterizing the physical. *Philosophical Studies*, 131(1): 61–99. DOI: https://doi.org/10.1007/s11098-006-5984-8

Wilson, J 2016 Grounding-based formulations of physicalism. *Topoi*, 37(3): 495–512. DOI: https://doi.org/10.1007/s11245-016-9435-7

Yablo, S 1992 Mental causation. *Philosophical Review*, 101(2): 245–280. DOI: https://doi.org/10.2307/2185535

Yablo, S 1993 Is conceivability a guide to possibility? *Philosophy and Phenomenological Research*, 53(1): 1–42. DOI: https://doi.org/10.2307/210805