



## The Oxford Handbook of the Philosophy of Consciousness

Uriah Kriegel (ed.)

<https://doi.org/10.1093/oxfordhb/9780198749677.001.0001>

Published: 2020

Online ISBN: 9780191813948

Print ISBN: 9780198749677

Search in this book

CHAPTER

# 1 The Problem of Consciousness

David Papineau

<https://doi.org/10.1093/oxfordhb/9780198749677.013.1> Pages 14–36

Published: 09 July 2020

## Abstract

Consciousness raises a range of philosophical questions. We can distinguish between the *How?*, *Where?*, and *What?* questions. First, *how* does consciousness relate to other features of reality? Second, *where* are conscious phenomena located in reality? And, third, *what* is the nature of consciousness? In line with much philosophical writing over the past fifty years, this chapter will focus mostly on the *How?* question. What makes the question difficult is the existence of an alleged explanatory gap between consciousness and physical processes. This chapter is devoted to the source of this explanatory gap and to the philosophical implications of a correct conception of this source.

**Keywords:** [Consciousness](#), [explanatory gap](#), [intuition of distinctness](#), [hard problem](#), [phenomenal concepts](#), [zombies](#)

**Subject:** [Philosophy of Mind](#), [Philosophy](#)

**Series:** [Oxford Handbooks](#)

**Collection:** [Oxford Handbooks Online](#)

## 1.1 Introduction

Consciousness raises a range of philosophical questions. We can distinguish between the *How?*, *Where?*, and *What?* questions. First, *how* does consciousness relate to other features of reality? Second, *where* are conscious phenomena located in reality? And, third, *what* is the nature of consciousness?

In line with much philosophical writing over the past fifty years, this chapter will focus mostly on the *How?* question. Towards the end I shall also say some things about the *Where?* question. As for the *What?* question, a few brief introductory remarks will have to suffice.

This is not to deny that the *What?* question raises a range of philosophically interesting issues. There is much to ask about the nature of consciousness. Must conscious states always involve some reflective awareness of themselves? Do all conscious states have an intentional content? Must consciousness always be consciousness for some subject, and if so for what kind of subject? In what sense, if any, are the conscious experiences of a subject at a given time always unified into some whole?

However, in order to keep my task manageable, I shall leave issues like these for the essays that follow in this volume. For my purposes, it will be enough if we simply characterize consciousness in the normal way, as states that are 'like something' for the subjects that have them. If examples are wanted, simply reflect on the difference between having your eyes open, and enjoying a range of conscious visual experiences, and closing your eyes and eliminating those conscious experiences. Or, more generally, contrast the conscious life you enjoy when you are awake, in all its rich variety, with its complete disappearance when you are given a general anaesthetic.

## 1.2 The Case for Physicalism

p. 15

Let us make a start on the *How?* question. How does consciousness relate to other features of reality? At first pass, it might be unclear why there is any special issue here. Why is there a puzzle about conscious states, as opposed to other kinds of states? Reality contains many different kind of things, biological, meteorological, chemical, electrical, and so on, all existing alongside each other, and all interacting causally in various ways. There seems no immediate reason why consciousness should be singled out as posing some special puzzle about its relation to the rest of reality.

If conscious properties did interact causally with non-conscious states, then there would indeed be no special problem about its relation to other features of reality. We could all be happy interactionists, in the style of Descartes. We could hold that conscious mind influences non-conscious matter, by controlling bodily behaviour, and similarly that matter influences mind, giving rise to sensory experiences, pains, and other conscious mental states.

There is a compelling argument, however, against this kind of interactionist stance. This derives from the so-called 'causal closure of the physical'. The problem is that the physical realm seems causally sufficient unto itself. Physical effects always issue from physical causes. This applies to bodily movements, and the neural processes which prompt them, as much as to any other physical effects. Scientists studying neural processes take it as given that the events they observe are effects of electromagnetic and chemical causes, not of independent mental influences exerting an influence from outside the physical realm.<sup>1</sup>

The 'causal closure of the physical' thus implies that, if there is a separate realm of mental states, it cannot exert any influence on bodily behaviour or other physical processes. One possible move at this point is to continue to uphold the existence of a distinct mental realm, and accept that it indeed has no influence on the physical world. However, this 'epiphenomenalist' option is not only intrinsically implausible, but faces various internal difficulties.<sup>2</sup>

Given this, most contemporary philosophers have opted instead for some form of physicalist monism. There are not two separate realms, mind and matter. Rather mental states are themselves a species of physical states. You might initially think of your pains or your desires as something separate from the cerebral and other physical states that accompany them. But in truth, so the physicalist thought goes, your mental states are one and the same as those physical states. On this view, of course, there is then no difficulty about pains, desires, and other mental states causally influencing bodily behaviour or other physical processes. If your conscious mental states are no different from your cerebral physical states, then they will have just the same physical effects that those physical states do.

p. 16

If there is such a compelling argument for a physicalist view of the mind, why hasn't physicalism always been the dominant philosophical position, rather than only becoming so in the middle of the last century? The answer is that the causal closure of the physical was not generally accepted until relatively recently. Note that the closure-based argument against dualism is not just the traditional objection put to Descartes by his contemporaries, that if mind and matter are so different, it is difficult to understand how they can causally interact. Even if this was a problem in Descartes' time, it is not clear that it greatly perturbed subsequent thinkers. Rather the problem is that modern science has definite views about the kinds of things that do affect the movement of matter, and independent mental influences are not among them.

To repeat, this exclusion of independent mental causes is a relatively recent phenomenon. Through most of the modern period, science had no problem with fundamental conscious causes. Orthodox physical science, from the time of Newton through to the twentieth century, was generally open-minded about the kinds of things that could influence the movements of matter. In addition to mechanical forces of impact, and gravitational forces, it allowed distinctive chemical forces, magnetic forces, forces of cohesion, vital forces—and conscious mental forces. 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.<sup>3</sup> It is noteworthy that all the familiar modern arguments for physicalism were developed in the middle of the twentieth century, and all appealed to some version of the causal closure of the physical.<sup>4</sup>

### 1.3 The Explanatory Gap

---

There is a huge contemporary literature on physicalist views of the mind, covering a range of questions. How exactly should we define 'physical'? Can mental properties be identified with basic physical properties, or should we instead embrace some version of 'non-reductive physicalism', according to which mental properties supervene on, or are grounded in, or are otherwise constituted by basic physical properties, without being strictly identical to them? Do these non-reductionist options succeed in avoiding the epiphenomenalist threat that prompted physicalism in the first place? And so on.

p. 17

However, we can by-pass all these issues here. This is because *any* version of physicalism about conscious states seems to generate pressing philosophical problems. Despite the strength of the argument for physicalism, the equation of the lived experience of perceptions, emotions, and pains with neuronal oscillations in the brain strikes many philosophers as effectively incomprehensible. As Thomas Nagel puts it in *The View from Nowhere* 'We have at present no conception of how a single event or thing could have both physical or phenomenological aspects, or how if it did they could be related' (1986: 47). Or, in the more direct words of Colin McGinn, 'How can technicolour phenomenology arise from soggy grey matter?' (1991: 1). To Nagel, McGinn, and many other philosophers, the idea that conscious states are at bottom physical seems obviously problematic.

This is the central problem of consciousness for contemporary philosophy. Arguments from causal closure provide compelling reasons to view conscious states as physical. But any such physicalist view of consciousness strikes many as little short of unintelligible. (Since this problem arises for all versions of physicalism, non-reductive as well as reductive, I shall often simplify the exposition from now on by talking of conscious states as physical, or as identical to physical states; everything that follows will apply equally to physicalist positions that view conscious states as supervenient on, or grounded in, or constituted by, physical states.)

If we are to make any progress with this central problem of consciousness, we need to articulate the nature of the resistance to physicalism illustrated by the quotations from Nagel and McGinn. One useful way to do this is to compare putative mind-brain identities with similar scientific identity claims in other areas. When

we are told that common salt is NaCl, or lightning is atmospheric electrical discharge, we happily accept these claims as telling us about the underlying physical nature of these everyday phenomena. But when we are told that pains are the firing of prefrontal nociceptive-specific neurons, or that visual experiences of red are neuronal oscillations in the V4 area of the visual cortex, we react quite differently. Even after we are given this information, we still want to know *why* those brain states are accompanied by those feelings. Why do the nociceptive-specific neurons, or the oscillations in V4, feel like that, rather than some other way, or no way at all? As Joseph Levine has put it, mind-brain identities seem to leave us with an *explanatory gap*, in a way that other scientific identities do not (Levine 1983). We remain puzzled about why the brain states give rise to the feelings, in a way that we do not feel puzzled about why NaCl gives rise to salt, or electrical discharges to lightning.

Now, as a social or psychological phenomenon, the existence of an explanatory gap is quite uncontentious. There is no doubt that most people react to mind-brain identity claims with demands for further explanation, in a way they don't react to other scientific identity claims. However, the philosophical significance of this social fact is far less straightforward. Philosophers disagree widely about the source of the reaction and about what, if anything, it implies about the relation between conscious and physical states.

p. 18 There are two distinct questions here. The first is a psycho-social question. What is the *source* of the explanatory asymmetry? Why do people feel that mind-brain identities, unlike scientific identities, leave something unexplained? The second is a philosophical question. What *follows* from this explanatory asymmetry? Does the puzzlement occasioned by mind-brain identities imply that there is some deficiency in the physicalist view of consciousness?

I shall address these issues in turn. The next three sections will be devoted to the source of the explanatory asymmetry. After that I shall turn to the philosophical implications.

## 1.4 The Derivability Gap

The psycho-social question first. My own view on this is a straightforward one. I think that the feeling of an explanatory gap is simply an upshot of the fact that we all—including professed physicalists like myself—find mind-brain identities almost impossible to believe. Even after we are shown plenty of evidence that pains and nociceptive-specific neuronal firing always accompany each other, and have the same causes and effects, we still intuitively resist the conclusion that they are identical. How could that urgent feeling possibly be one and the same as neuronal activity, we ask? We find it hard to escape the spontaneous dualist thought that the feeling and the physical state are not one thing, but two different states that somehow invariably accompany each other.

And, to the extent that we do think this, then of course we feel a need for more explanation. Why is the neuronal activity accompanied by the nasty feeling of pain, rather than a pleasant sense of floating, say? Indeed, why is it accompanied by any feeling at all? Once we slip into the dualist way of seeing things, we cannot avoid a range of demands for further explanation. (See Papineau 2010.)

Most philosophers currently working on consciousness, however, take a quite different and less straightforward view of the feeling of an explanatory gap. In their view, this feeling is not a consequence of an intuitive resistance to physicalism. Rather, it stems from an internal feature of the way we think of conscious states, and would persist even if we were able fully to embrace physicalism.

This mainstream view attributes the feeling of an explanatory gap to the impossibility of *deriving* mind-brain identities a priori from the physical facts. This is supposed to mark a contrast with the scientific cases.

While we can often derive scientific identities a priori from the physical facts, so the thought goes, we cannot so derive mind-brain identities, and this creates a feeling of puzzlement about them.

p. 19 The reason for the difference, on this mainstream account, lies in the different ways in which we ordinarily conceive of scientific properties and conscious properties. Consider our everyday concept of *common salt*. According to the mainstream view, we think of salt as the stuff that is white, crystalline, granular, with a distinctive taste, that dissolves in water, and is found in the oceans. Now imagine someone who has a fully detailed ↵ account of the physical make-up of the world, in terms of the distribution of matter, arrangement of elementary particles, the deployment of fields, and so on. In principle, such a person could arguably put this knowledge together with their prior conceptual grasp of *salt* to figure out that salt must be NaCl, on the grounds that NaCl is the stuff that fits the conceptual requirements for salt—white, crystalline, ... However, we can't do this with pain, say, or with visual experiences of red. The problem is that our everyday concepts of *pain* or *visual experience* do not pick out their objects via some descriptive role, like white, crystalline, ... but in terms, so to speak, of what the states feel like. We typically think of conscious states directly, by focusing on the feelings involved, and not as the states, whatever they may be, that play some descriptively specified role. And this blocks any a priori derivation of mind-brain identities from the physical facts, of the kind that is arguably available for identities like salt = NaCl. Scrutinize the physical facts as much as you like, and they won't tell you that pains are the firing of prefrontal nociceptive-specific neurons. Since we don't think of pains in terms of some specified role, but in terms of the feelings involved, there is no way to connect the physical facts with the phenomenon of pain.

Given this, our knowledge of mind-brain identities can only be based on some kind of a posteriori abductive inference, rather than a principled a priori demonstration that a certain physical state fills some specified role. For example, we might observe that pains occur whenever prefrontal nociceptive-specific neurons fire, and vice versa; we might also note that, if pains were the firing of nociceptive-specific neurons, then this would account for a number of other observed facts about pain, such as that it can be caused by trapped nerves, and can be blocked by aspirin; and we might conclude on this basis that pains are indeed identical to the firing of nociceptive-specific neurons.

Still, to repeat, there is no question of deriving this identity a priori from the physical facts, by showing that the nociceptive-specific neuronal firing fills the pain role—for we do not think of conscious pains in terms of roles to start with. And this, says the view under examination, is why we feel an explanatory gap in the mind-brain case. The peculiarly direct nature of our concepts of conscious states stops us deriving mind-brain identities a priori from the physical facts.

## 1.5 Doubts about Derivability

---

This lack-of-derivability account of the source of explanatory-gap feelings is widely taken for granted in contemporary philosophy of mind. Despite this, I think it is clearly mistaken, and shall explain why in a moment. One part of the story, however, is relatively uncontroversial. This is the idea that we have direct, non-descriptive concepts of conscious states that preclude any a priori derivation of mind-brain identities from physics.

p. 20 Some initial mid-twentieth-century versions of physicalism did not accept this, and so held that mind-brain identities could indeed be read off from the physical facts. ↵ But this stance was dealt a critical blow by Frank Jackson's 'Knowledge Argument' (1986). Jackson pointed out that someone who has never experienced colours could be in possession of all the physical facts about colour vision, and yet not 'know what it is like' to see something red. The overall philosophical significance of Jackson's argument is a complex matter, to which we shall return in due course. But it is pretty much agreed on all sides that, at a minimum, Jackson's argument does demonstrate the existence of a special range of 'phenomenal

concepts', ways of thinking about conscious states directly, in terms of the feelings involved, which are normally only available to subjects who have experienced those states themselves, and which block any physics-based derivation of mind-brain identities.<sup>5</sup>

So far so good. We cannot derive mind-brain identities a priori from the physical facts. Still, is this really the source of the feeling that the identities leave something unexplained, as claimed by the suggestion currently under examination? This suggestion faces an obvious objection. Plenty of other identities similarly cannot be derived from physics, but generate no corresponding impression of an explanatory gap.

After all, as the above remarks make clear, a priori derivations from physics will be blocked whenever we have concepts that refer directly, rather than by association with some described role. On the face of things, phenomenal concepts are by no means the only such cases. Proper names ('Cary Grant'), demonstrative constructions ('that dog'), and simple terms for observable properties of objects ('round') are all arguably terms that refer directly, rather than by description. Given this, when we accept identity claims involving these terms (such as 'Cary Grant = Archie Leach', or 'that dog = her pet', or 'round = locus of constant distance from some point'), it can only be on the basis of an abductive inference from direct empirical evidence, such as that the two things in question are found in the same places and the same times, and are observed to bear the same relations to other things, not because we can deduce the identities a priori from the physical facts.

Yet we feel no explanatory disquiet when presented with these identities. Even though they must perforce be based on some form of abductive inference, for lack of any descriptive roles associated with the relevant terms, they certainly do not leave us with a feeling that something has been left unexplained.

p. 21 Come to think of it, it is doubtful that many scientific identities are based on anything more than abductive inferences either. When nineteenth-century scientists first figured out that salt is NaCl, they certainly did not do so by inferring a priori from basic physical theory that NaCl molecules will appear white, form crystals, dissolve in water, and so on, and hence concluding that NaCl must be the substance that fits the specifications for salt. The sub-atomic understanding required for such derivations was more than a century in the future. Rather the original scientists simply noted that NaCl molecules were present whenever salt was, and vice versa, and had some of the same causes and effects, and identified them on that basis.<sup>6</sup>

The same goes for the identification of lightning with electrical discharge, or consumption with tuberculosis infection, or nearly all other scientific identities. Scientists did not derive these identities a priori from physical theory, but based them largely on simple observations of co-occurrence and matching casual relations to other things.

Yet this did not make the scientists feel something was left unexplained. Even though the identities were based on abductive inference, rather than derived a priori from the physical facts, the scientists were not left unhappily puzzled about why NaCl gives rise to salt, rather than to something else, or to nothing at all, or why lightning arises from electrical discharge, or consumption from tuberculosis infection. Once more, the absence of a priori derivations did not seem to engender any feelings of things left unexplained.

Perhaps these counterexamples are less than conclusive. Still, there are further grounds, apart from possible counterexamples, for doubting the mainstream thesis that the feeling of an explanatory gap arises from our inability to derive the relevant identities a priori from the physical facts.

Note how this mainstream account implies that *something is left unexplained* when we embrace mind-identities on the basis of abductive inferences, something that *does get explained* when we supposedly derive scientific identities from the physical facts. But what exactly is that? What exactly does get explained,



according to the mainstream account, when we derive scientific identities a priori from the physical facts, but not when we embrace mind-brain identities on the basis of abductive inferences?

One first thought might be that it is the identities themselves that are explained. We can explain *why* salt is NaCl, or *why* lightning is electrical discharge, once we can derive the identities a priori from the basic physical facts, in a way these identities would be left unexplained if we simply based them on abductive inferences.

But this seems odd. We do not normally regard identities as in need of explanation. Since they are necessary, they could not have been otherwise, and did not need anything to make them so. (To repeat a familiar example, when we discover that Mark Twain = Samuel Clemens, we might reasonably ask *why* he had two names, or *why* nobody told us before. But it would make no sense to ask—*why* was Mark Twain the same man as Samuel Clemens? (Block 1978.))

A second thought would be that it is not the identities per se that get explained, but the behaviour displayed by everyday kinds. So, for example, once we can derive the identity of salt with NaCl a priori from the physical facts, then we will in principle be able to explain why salt displays such defining characteristics as whiteness, crystallinity, solubility in water, and so on, in a way we can not do if the identity is based on brute correlations.

p. 22 But this second thought does not hold water either. Sure, we can explain the behaviour of salt if we derive its identity with NaCl a priori from the physical facts, by appealing to our physical understanding of how NaCl molecules work. But, by just the same coin, we can explain the behaviour of pain if we accept its identity with nociceptive neuronal firings on the basis of an a posteriori abductive inference. As I observed above, an identification of pain with the firing of prefrontal nociceptive-specific neurons, even if based on an abductive inference, will happily allow us to explain such things as why pain is caused by trapped nerves, or why it is relieved by aspirin.

So, once more, it does not look like the impression of an explanatory gap can really be due to a lack of a priori derivation. Such derivations do not allow us to explain anything that cannot be explained without them.

Much of the contemporary literature on the 'explanatory gap' simply reads this phrase as referring to the impossibility of deriving mind-brain identities a priori from the physical facts. But we have now seen that, in truth, this understanding quite fails to answer the psycho-social question of why mind-brain identities leave most people with the feeling of an *explanatory* gap. For a start, people do not seem to have any feeling of non-explanation with other identities that cannot be derived a priori from the physical facts. Moreover, nothing extra would seem to be explained when we can derive identities a priori from the physical facts.

In the end, though, there is an even more powerful reason for rejecting the idea that the feeling of an explanatory gap is something to do with a priori underivability. This is the availability of the alternative account mentioned earlier, an account that avoids all the difficulties raised in this section.

## 1.6 The Intuition of Distinctness

On the alternative account I favour, the issue is not that we feel that something still remains to be explained *after* we have accepted mind-brain identities. It is rather that we all find mind-brain identities very difficult to accept in the first place.

As I observed above, even after we are given all the abductive evidence, we still find mind-brain identity claims almost impossible to believe. We cannot resist the dualist conviction that conscious feelings and the

physical brain states are two different things. And this, in my view, is the real reason why we feel a need for further explanation. We want to know why the neuronal activity is accompanied by *that* conscious feeling, rather than by some other, or by no feeling at all. Our dualist intuitions automatically generate a hankering for further explanation.

On my diagnosis, then, the demand for explanation arises, not because something is lacking in physicalism, but because something is lacking in us. Even after we are shown the arguments for physicalism, and are perhaps moved to embrace physicalism at a theoretical level, we continue to experience the pull of the dualistic perspective, and so intuitively feel that something remains to be explained.

p. 23 If only we could fully embrace physicalism, this diagnosis suggests, the feeling of an explanatory gap would disappear. If we could fully accept that pains are nociceptive-specific  $\hookrightarrow$  neuronal firings, then we would stop asking why 'they' go together—after all, nothing can possibly come apart from itself. The feeling of a gap is simply a corollary of the intuitive grip of dualism.

From this perspective, then, a properly thorough-going physicalism promises to dissolve 'the problem of consciousness'. The committed physicalist will simply deny that any puzzle is raised by the fact that it feels painful to be a human with active nociceptive-neurons. Why shouldn't it feel like that? That's how it turns out. Why regard this as puzzling?

Note how my diagnosis in terms of intuitive dualism offers a far better account of the feeling of an *explanatory* gap than the appeal to lack of a priori derivability. For a start, it is now clear why we feel something has been left *unexplained*—we want to know specifically why brain states give rise to extra conscious states. Moreover, the feeling of a gap is now specifically about mind-brain relations, and so there is no puzzle about why we do not feel it in other cases where a priori derivability is blocked.

By way of further support for the idea that the feeling of an explanatory gap stems from intuitive dualism, we need only attend to the phraseology normally used to discuss the relation between mind and brain. Brain processes are said to 'generate', or 'yield', or 'cause', or 'give rise to' conscious states. ('How can technicolour phenomenology *arise from* soggy grey matter?') These expressions are common currency in writings on consciousness, including by thinkers who say they are not dualists. But the phraseology itself is not consistent with physicalism. Fire 'generates', 'causes', 'yields', or 'gives rise to' smoke. But NaCl does not 'generate', 'cause', 'yield', or 'give rise to' salt. It is salt. The point is clear. To speak of brain processes as 'generating' conscious states, and so on, only makes sense if you are implicitly thinking of the conscious states as separate from the brain states.

If yet further evidence is needed, consider our intuitive reaction to whether zombies are possible. Could a being share all your physical properties but have no conscious life? Everybody's first thought is, 'Sure. Just duplicate the physical stuff and leave out the feelings.'

Reflective physicalists will of course realize, on second thoughts, that they must deny that this is really possible. (If conscious states *are* physical states, the 'two' cannot come apart.) But it is the initial reaction that I want to focus on here. Compare our response to the idea of Marilyn Monroe existing without Norma Jean Baker, say. I take it that our initial reaction to this suggested possibility would be puzzlement. What are we being asked to posit? That she exist without herself? That makes no sense.

This contrast is a reflection of our intuitive dualism. Zombies strike us as initially possible simply because all of us, physicalists included, intuitively think of conscious feelings and physical states as distinct existents. If we fully embraced the idea that they are one and the same, then we would find the idea of zombies simply puzzling. How could there be nociceptive-specific neuronal firing without pains? What are we being asked to posit? That the state exist without itself? That would not make any more sense than Marilyn Monroe without Norma Jean Baker. (Cf Papineau 2007.)



If the feeling of an explanatory gap stems from our intuitive dualism, as I have been arguing, then the obvious next question is about the cause of these persistent dualist thoughts. Why do dualist ideas maintain such a firm grip, even on thinkers who are fully persuaded of the strength of the arguments for physicalism?

Plenty of possible answers to this question offer themselves, but before considering them I would like first to return to the issue left hanging earlier, namely whether the feeling of an explanatory gap is associated with any good arguments against a physicalist view of consciousness. After all, one possible explanation for why many people feel intuitively convinced that physicalism is false might be that they can all see that there is a strong argument against it.

Of course, even if there were a good philosophical argument against physicalism, it might not be the reason most people instinctively reject physicalism; the argument might not be apparent to them. But, even so, it will be useful to get clear about the nature of the arguments against physicalism, before discussing the possible causes of persistent dualist intuitions.

## 1.7 Arguments Against Physicalism

The best place to begin assessing the argumentative case against physicalism is with Jackson's 'Knowledge Argument'.

As I explained earlier, Jackson's argument hinges on the observation that someone could know all the physical facts about colour vision, and yet not 'know what it is like' to see something red. And, as I said, this observation is generally agreed to demonstrate the existence of a special range of 'phenomenal concepts' that refer directly to conscious states and are normally only available to subjects who have experienced those states themselves.

Jackson's original intention, however, was not just to argue for an extra set of phenomenal concepts, but in addition for an extra set of phenomenal properties. That is, he was arguing for the dualist conclusion that conscious states are metaphysically distinct from and additional to any physical states.

Even so, once phenomenal concepts are on the table, physicalists would seem to have a ready initial response to his argument. They can say that 'not knowing what some conscious states are like', even when you are completely knowledgeable about the physical facts, is simply a matter of not being able to represent certain physical states (the relevant conscious states) using phenomenal concepts. Once you know all the physical facts, then you know about all of reality. If, despite this, you still 'don't know what some states are like', that is just a matter of your not being able to represent those states in the special direct way that only becomes available once you are in possession of the relevant phenomenal concepts.

From the point of view of physicalists who take this line—'a posteriori' physicalists—there are thus two distinct kinds of concepts that refer to conscious states. On the one hand are phenomenal concepts—like *pain* or *seeing something red*—that pick out their referents directly, in terms of what they feel like, so to speak. And on the other are physical concepts—like *nociceptive-specific neuronal firing* or *oscillations in V4*—that refer to just the same states in terms of their physical nature. Scientific investigation can then show us that the former concepts pick out the same things as the latter ones, just as it establishes such other a posteriori identities as *salt = NaCl*, or *lightning = atmospheric electrical discharge*.

However, a second line of anti-physicalist argument now comes into play. This focuses on the particular nature of phenomenal concepts, and contends that certain features of these concepts are incompatible with their referring to physical states.

The basic thought is that, if physicalism were true, the *directness* of phenomenal concepts ought to render its falsity inconceivable—yet it does not. Consider once more a ‘zombie’, a being who shares all my physical properties yet has no conscious life. Physicalists must deny that zombies are *possible*, given that the mind is ontologically inseparable from the brain. But a posteriori physicalists have no choice but to allow that they are at least *conceivable*. (If phenomenal concepts refer directly, and my feelings are therefore not a priori derivable from my physical properties, then there is no conceptual contradiction in ascribing a being all my physical properties, but denying it my conscious ones.)

The argument against physicalism now hinges on the thesis that impossibilities are only conceivable when presented using concepts that refer *indirectly*. For example, take *salt* = NaCl. Even though this could not be otherwise—salt is NaCl—someone can certainly conceive of (indeed believe in) NaCl not being salt. However, according to the argument at issue, they can only do this because they are thinking of salt at second hand, as *the* substance, whatever it is, that is white, crystalline, ... This way of thinking leaves it open whether or not salt is in fact NaCl, and thus whether it is necessarily identical to that substance.

But a phenomenal concept like *pain* is not indirect in this way. Phenomenal concepts do not pick out their referents indirectly, by some association with a role, as with *salt*, but directly, in terms of what they are like. So there is no room, the argument goes, for claims made using phenomenal concepts, such as *nociceptive-specific firing is pain*, to be necessarily true, yet appear conceivably false. If this claim were true, it would have to be a priori. Yet it is not.

The crucial premise in this argument is that necessary facts can only appear conceivably false when they are presented using indirect concepts. Or, putting it the other way around, once everything is formulated directly, no necessary truths will appear conceivably false. In short, the crucial premise is that direct concepts are *revelatory*, in the sense of displaying all the necessary properties of their referents a priori. (And then the anti-physicalist reasoning goes: the concept *pain* is direct, so it must be revelatory; so, if my physical nature necessitated my pains, this ought to be knowable a priori; but it is not; therefore pains cannot be physical.)

A posteriori physicalists deny the crucial premise of this argument. They do not accept that direct concepts are always revelatory. Directness is a semantic matter—the concept picks out its reference directly, rather than as the item that satisfies some descriptive role. Revelatoriness is epistemological—the concept renders all necessary features of its referent a priori knowable. A posteriori physicalists insist that the former does not imply the latter.

p. 26

In particular, they hold that phenomenal concepts are direct but not revelatory. They accept that phenomenal concepts are direct. And as physicalists they of course hold that pains have a physical nature. But they deny that this essential feature of pains must be revealed to us by the phenomenal concept *pain*. You can grasp this direct concept fully, yet not appreciate that pains are necessitated by the relevant brain processes.

An extensive literature is devoted to the question of whether all direct concepts are revelatory, and all directly formulated necessary claims are therefore knowable a priori. (See e.g., Block and Stalnaker 1999; Chalmers and Jackson 2001; Chalmers 2002; Levine 2001, 2010.) A posteriori physicalists and other opponents of this thesis contend that there are plenty of counterexamples. What about identity claims involving proper names, indexical constructions, or observational concepts—‘Cary Grant = Archie Leach’, ‘that dog = her pet’, ‘round = locus of constant distance from some point’? Earlier I cited the apparent a priori underderivability of such claims from the physical facts as an argument against attributing the explanatory gap to this kind of underderivability. In the present context, the same cases offer putative examples of directly formulated necessities that are not a priori knowable.

At this point the arguments get messy. Anti-physicalists respond that, despite the prima facie absence of descriptive content, the terms in question should properly be understood as functioning indirectly, and that this is why they do not reveal the identity claims involving them a priori. Some physicalists counter by questioning the way their opponents are drawing the distinction between direct and indirect terms. An alternative physicalist strategy is to grant that in general direct concepts are revelatory and directly formulated necessities a priori, and that a posteriori phenomenal mind-brain claims are therefore an exception to this rule, but maintain that there is nothing wrong with that. And so on. (See Levine 2001: ch 2.4.)

Fortunately, it is possible to cut through much of this dialectic. What really matters for the anti-physicalist argument is whether *phenomenal concepts* are revelatory, not any more general thesis about some wider category of 'direct' concepts. The anti-physicalists say that phenomenal concepts are revelatory, and in particular that they reveal conscious states not to be physical. Physicalists respond that there is no reason to suppose that phenomenal concepts have the power to reveal such things.

Given this, it makes sense for us to address the revelatoriness of phenomenal concepts head on, and by-pass the further issue of whether this can be seen as a special case of some more general principle involving direct concepts. As far as the anti-physicalist argument goes, all that matters is the workings of phenomenal concepts themselves. (Cf Nida-Rümelin 2007; Goff 2011.)

At an intuitive level, it is certainly not implausible that phenomenal concepts are revelatory. Consider what it is like to think at first hand about a stabbing pain, or a visual experience of seeing something red. Does not such thinking acquaint you with the very nature of these conscious states? It certainly seems as if such phenomenal thinking lays bare all essential aspects of the relevant experiences.

p. 27 A posteriori physicalists will respond that appearances are deceptive. We should not be distracted, they will say, by the close association between phenomenal thinking and the experiences being thought about. Often the experience itself (the pain, the awareness of red) is present when we think about it phenomenally. In other cases, an imagined version of the experience (a 'faint' copy', as Hume put it) accompanies our phenomenal thinking. And, because of this, it can seem that everything is revealed. A version of the experience is right there, before our minds. How can anything essential remain hidden?

But it is one thing, physicalists will object, to *have* an experience. It is another to *know* everything about its nature. Phenomenal thinking might characteristically *give* us the experience, in the sense that we undergo some version of it while thinking about it. But this does not mean it tells us everything about its nature. In particular, it does not mean it will reveal that the experiences are at bottom physical, if they are.

Moreover, the physicalist can continue, there is something deeply mysterious about the idea that merely thinking about something can reveal all its necessary properties. Of course, in the case of complex concepts with internal structure, mere thinking can deliver analytic knowledge; for example, someone who possesses complex concept *square* can work out, just by analysing this concept, that *squares have four sides*. But this model does not seem relevant to the putative power of phenomenal thinking. Phenomenal concepts like *pain* or *seeing something red* do not seem complex; nor, correspondingly, do anti-physicalists maintain that the non-physicality of their referents is an analytic consequence of their internal structure.

Perhaps anti-physicalists can appeal to a different model. Instead of invoking analytic knowledge, they can construe phenomenal thinking as a kind of direct acquaintance, appealing to the point that such thinking is characteristically accompanied by versions of the experiences thought of. The idea would be that we find out about phenomenal states by thinking about them introspectively. We scrutinize our experiences internally, and thereby uncover their nature.

But the mystery remains. Introspection is certainly able to tell us what experiences we are having, and various other things about them. But why should it be guaranteed to tell us about *all* their necessary properties? How is that supposed to work? Any normal information-delivering process is inevitably fallible and only partially informative about the nature of its objects. To hold that introspection is guaranteed to reveal all necessary properties of experience would seem to take us beyond the realm of naturally explicable faculties.

## 1.8 Neutral Monism

---

p. 28 Suppose for the moment that the argument from revelation did hold water. This would scarcely leave the anti-physicalist in a comfortable position. As I observed earlier, modern scientific findings seem to leave epiphenomenalism as the only viable alternative ↯ to physicalism. Yet the epiphenomenalist relegation of conscious states to inefficacious causal ‘danglers’ is not an attractive option. If this is where the argument from revelation ends up, that would itself be a reason for thinking it must have gone wrong somewhere.

But perhaps there is another way out. An increasing number of contemporary philosophers favour an alternative view, known as ‘Russellian monism’, which offers a way of embracing the argument from revelation while avoiding the entanglements of epiphenomenalism. In effect, this position aims to maintain the causal significance of phenomenal states by viewing both the phenomenal and the physical as grounded in some more fundamental reality.

Let us go back to the argument from revelation. This said that a truth of the form *pains = nociceptive-specific firing* can only be conceivably false if it is formulated in indirect terms. The route from this to Russellian monism hinges on the thought that perhaps it is *nociceptive-specific firing* that is the indirect term, rather than *pain*.

So far I have not queried the idea that physical terms like *nociceptive-specific firing/NaCl/electric discharge* are direct and revelatory. But there is no reason to take this for granted. A standard account of scientific terms has them referring via theoretical descriptions—to that property, or quantity, that plays such-and-such a theoretically specified role. (So for example, *mass* might be equated with *that quantity that is inversely proportional to acceleration and obeys the law of gravitation*.)

This now offers a different way of squaring the conceivable falsity of *pains = nociceptive-specific firing* with the principle that necessary truths can only be conceivably false if formulated in indirect terms. Suppose that the term *nociceptive-specific firing* refers indirectly to that underlying property, whatever it is, that plays the role specified by neurophysiological theory. Russellian monism now views the conscious feeling of pain as itself grounded in this underlying property (Russell 1927).

This allows us to account for the conceivability of zombies, beings who have nociceptive-specific firing but no pains, as possible beings in whom the relevant theoretical role is filled, not by the underlying property that constitutes pain in the actual world, but by some different and non-conscious property. Since we are thinking of the nociceptive-specific firing only indirectly, as the filler of a theoretically specified role, this leaves it open that this role could possibly be played by something other than its actual filler, indeed by something that fails to constitute any conscious feeling at all.

At the same time, this Russellian move promises to eliminate any worries about the epiphenomenality of pain. After all, pain is now constituted by a basic property, the property that fills the *nociceptive-specific firing* role in the actual world. At first pass, such basic properties look like just the kind of items to enter into fundamental causal relations.

p. 29

This Russellian position is often associated with some version of the *panpsychist* doctrine that consciousness permeates all parts of the natural world. For some thinkers, this further commitment is motivated by the thought that our introspective awareness of our conscious experience is the only point at which we are directly acquainted with the underlying nature of reality. Since introspection shows reality to be conscious in all  $\hookleftarrow$  cases where its underlying nature is revealed, the thought continues, we should therefore conclude that it is conscious throughout (Goff 2017).

A further motivation for panspsychism derives from a perceived need to *explain* the consciousness that is present in beings with brains like ours. Russellian monists accept the orthodox view that the underlying physical processes that constitute our conscious life are complex, and in particular that they are built up from the same simple components (fundamental field and particles) that compose the rest of nature. Given this, many feel that it would be mysterious for consciousness to emerge in complex brain processes if it were not already present in the simple parts. (Cf Strawson 2003.)

Despite Russellian monism's current popularity, it is questionable whether it marks any real advance on ordinary a posteriori physicalism. On further analysis, it turns out to leave us with many of the same issues, and moreover to generate a number of problems of its own.

An initial difficulty relates to the explanation of macroscopic conscious states in terms of their microscopic parts. Even if the microscopic components are credited with some conscious nature, this will presumably be different in kind from the conscious nature of the wholes they compose. So why is the relation between the conscious parts and the differently conscious wholes any less mysterious than the supposedly puzzling emergence of conscious wholes from non-conscious parts? (Stoljar 2006.)

A converse puzzle involves our phenomenal knowledge of macroscopic conscious states with microscopic parts. If phenomenal concepts reveal all the necessary properties of their referents, then why do they not show pains and other conscious states to be composite? If some state is built from parts, then this is presumably part of its nature. Yet introspection presents conscious states like pains as simple and unified, not composite (Lockwood 1993).

A further worry is that Russellian monism seems to end up flirting with the very epiphenomenalism it is designed to avoid. It is essential to the Russellian position that the *nociceptive-specific firing* role, say, might possibly be filled by a number of different underlying states, including ones that have no conscious nature (as in the zombie version of me). But now it looks as if the conscious differences between these alternative fillers make no difference to their causal powers. After all, by hypothesis these different fillers all display just the same behaviour and conform to just the same scientific laws. If, in addition, the fillers involve variations in consciousness, these variations would thus seem condemned to causal inertness.<sup>7</sup>

p. 30

Finally, and relatedly, the general metaphysical position on which Russellian monism rests is itself highly contentious. As the Russellians see it, scientific terms are non-revelatory because the specification of a theoretical role leaves it open which underlying entity fills that role. But it is not obvious, to say the least, that we should accept this thesis. Consider the case of mass. As I said, science arguably picks this out as *that quantity that is inversely proportional to acceleration and obeys the law of gravitation*. From the  $\hookleftarrow$  Russellian perspective, then, there is another possible world, just like the actual world, save that some different quantity, *schmass*, plays the mass role there. But this seems a perverse commitment. Surely that would simply be another world that contains *mass*, the same quantity as is present in our world.

This is not the place to resolve the debate about the metaphysical relation between properties and laws (cf Bird 2007). Still, on the face of things, the more natural view would seem to be that basic scientific properties are necessarily attached to their nomological roles. Fix the profile of laws that governs the entity, and you have fixed the entity itself.<sup>8</sup> Why multiply complexity unnecessarily by positing differences that have no further consequences?

All in all, then, Russellian monism seems to generate more problems than it solves. In my view, we would do better to stick with simple a posteriori physicalism, and forget about the supposed argument from revelation. Abductive evidence establishes certain phenomenal–physical identities. Even if both the phenomenal and physical concepts involved pick out their referents directly, the conceivable falsity of these identities does not discredit them. Why ever should we suppose that directly referring terms will reveal all the necessary features of their referents a priori?

## 1.9 Explaining the Intuition of Distinctness

---

Let us return to the feeling of an ‘explanatory gap’. My earlier diagnosis was that this is simply a manifestation of a widespread intuitive conviction that dualism is true. Even those who take themselves to be persuaded of physicalism cannot shake off the intuition that consciousness is non–physical, and so find themselves hankering for some explanation of how the brain ‘gives rise’ to conscious feelings.

This diagnosis then left us with a different explanatory need. Why does the intuition of dualism exert such a grip on our minds? One possibility was that we are all persuaded by our awareness of a sound argument for dualism. If there is no such argument,<sup>9</sup> however, then the explanation for the intuition of dualism must lie elsewhere.

p. 31 As it happens, there is no shortage of existing hypotheses about dualist intuitions. In a moment I shall describe some of these suggestions.

But first it is worth observing that these hypotheses are not in competition. A number of different factors may work together in persuading us against physicalism. Perhaps this itself is a large part of the reason dualist thinking is so persistent. The different psychological pressures favouring dualism gain strength by acting in concert. For each dualist influence that is identified and resisted, others are waiting in the wings, ready to capture our thinking again.

I shall now briefly run through six different theories that have been put forward to account for the prevalence of dualist thinking.<sup>10</sup>

*Culture.* Perhaps the widespread commitment to dualism is simply a reflection of the influence of religious metaphysics on much everyday thinking. The inherited culture of most societies is shaped by religious beliefs in non–physical realities. It might not be easy to shed the metaphysical ideas we acquire early on from such influences.

*Natural–Born Dualists.* Many anthropologists and developmental psychologists attribute cultural dualism, including the dualism of religious beliefs, to a more fundamental psychological source. As they see things, the very structure of human thinking inclines us to what they call ‘intuitive dualism’. From infancy onwards, the human mind automatically categorizes everything it sees as *either* active self–moving entities directed by minds *or* as passive physical processes triggered by external influences. This categorization thus excludes the possibility of anything that is simultaneously both mental and physical.

*Cognitive Architecture.* According to many philosophers of mind, when people come to believe an identity of the form  $a = b$ , they typically ‘merge their files’ for  $a$  and  $b$ . That is, they reorganize their cognitive architecture so that all the items of information previously associated with  $a$  and  $b$  respectively are now unified in a single  $a/b$  file. Perhaps, however, something blocks this merging in the special case of phenomenal–physical identities. Maybe phenomenal concepts are housed in the sensory cortex, while physical concepts are associated with linguistic areas of the brain. This might prevent us from performing the usual ‘merge’ operation with phenomenal–physical identities, and consequently make us feel that something is amiss with the identities themselves.



p. 32

*The Antipathetic Fallacy.* Recall a point made earlier, that phenomenal thinking about a conscious state is typically accompanied by some version of the state itself: when we think phenomenally about pain, we normally either have a pain, or we recreate a ‘faint copy’ in imagination. Now consider a thought like *pains = nociceptive-specific firing*. When we contemplate this thought, we might naturally enough form the impression that the pain itself, the feeling, is present on the left hand side—given that a version of the pain itself is likely to accompany this phenomenal thinking—whereas by contrast there is no such feeling on the right hand side—exercising the ↪ physical concept *nociceptive-specific firing* does not itself generally activate any pain. And we might for this reason conclude that the right hand side ‘leaves out’ the pain, and so does not succeed in referring to it. Now, of course this is a fallacy (I have dubbed it ‘the antipathetic fallacy’, Papineau 1993): even if the concept on right hand side does not *use* the pain, as the left hand side does, this is no reason to conclude that it does not *mention* it. But for all that, it is still a highly seductive fallacy.

*The Phenomenological Fallacy.* Everyday thinking (along with some philosophical theories) takes the view that sensory experiences are constituted in part by ordinary worldly properties: when you see a green circle, the properties of greenness and roundness are in some sense literally present in your experience. However, it is clear that these worldly properties are not instantiated in the brain—nothing in the brain is green or round. So the natural inference is that the sensory experiences must be distinct from brain processes. U. T. Place called this inference ‘the phenomenological fallacy’ and located the mistake in the initial premise that worldly properties are constituents of experiences (Place 1956). Perhaps Place was too quick to diagnose a fallacy. Maybe there are good senses in which worldly properties are constituents of sensory experience, and indeed senses in which this is consistent with physicalism. Still, however these niceties work out, the point remains that the manifest absence of properties like greenness and roundness from the brain might be the reason why many people are convinced that sensory experiences cannot be physical.

*Revelation.* Finally, recall the suggestion that phenomenal concepts are revelatory. In an earlier section I argued again this suggestion. But, as I said at the time, it is a highly intuitive idea. When you think at first hand about a stabbing pain, or visual experience of seeing something red, it certainly seems as if you are acquainted with the very nature of these conscious states, and that this reveals them to be non-physical. In my view, of course, introspection lacks the power to show us that experiences are non-physical. But I have no doubt that many people think it does show this, and embrace dualism for that reason.

## 1.10 Final Thoughts on the ‘Where’ Question

I have argued that a clear-headed physicalism resolves the ‘how’ question about consciousness. Conscious states are simply one and the same as physical states. The supposed ‘explanatory gap’ between brain and mind is nothing but a corollary of dualist intuitions. Certainly some physical states are like something for the beings that have them. But what is so surprising about that? That’s how it is.

p. 33

I would now like to conclude by briefly raising some doubts about the ‘where’ question. Much contemporary research is concerned with the location of conscious phenomena. Such research aims to discover which neural processes are like something for their ↪ subjects, and which are not. For instance, is early visual processing in cortical area V2 conscious for humans? Are neural processes in fish conscious? What about activity in insect brains?

Queries like these are the focus of a great deal of contemporary debate. But I am not sure that they are good questions. It seems to me possible that they too are a misplaced consequence of intuitive dualist commitments.

Ned Block has distinguished *phenomenal* from *access* consciousness (Block 1995). A state is phenomenally conscious if it is like something. It is access conscious if the subject can make use of it for reasoning and control of action. In beings who are capable of introspectively reporting their own mental states, the accessible conscious states will thus be the ones that normal subjects can report.

Access consciousness is certainly an interesting and significant category. Prior to the relevant research, who would have thought that people cannot always report the dorsal stream information that guides their hands' grasping movements (Milner and Goodale 2008), or that patients given morphine remain aware of their pains even after they cease to be distressing? We certainly want to know which cerebral states are accessible to subjects, and which not.

Still, given that we have this distinction, do we need any further division between states that are and are not *phenomenally conscious*? Why suppose that this concept draws a significant line in nature, now that we have taken care to distinguish it from the clearly-defined cognitive role of accessibly in the sense of contributing to reasoning and the control of action?

Of course, if phenomenal consciousness were constituted by some extra mind stuff, something additional to the physical realm, then there would be a real difference between the presence and absence of this mind stuff. However, once we free ourselves from the intuitive myth of such extra mind stuff, should we continue to think of phenomenal consciousness as constituting a distinctive physical kind?

I don't entirely want to rule out this possibility. By way of comparison, the notion of *life* used to be associated with the idea of an *élan vital*, some non-physical substance that animates living beings. But, even though we now reject any such non-physical substance, we still recognize animate beings as a significant sub-category of physical systems.

I doubt, however, whether an analogous point applies to consciousness. In the case of life, we can point to a well-defined range of features that make it worth differentiating living beings: self-sustaining, anti-entropic, reproductive. It is not clear that anything similar gives us a hold on phenomenal consciousness. We can of course distinguish some cerebral states as those to which we have introspective access. But it is not clear that, beyond that, we have any clear ideas about what distinctive features distinguish states that are phenomenally 'like something' from other physical processes.<sup>11</sup>

p. 34 Isn't the reality of phenomenal consciousness simply manifest? What about the technicolour phenomenology of visual experience, or the vicious unpleasantness of intense pain? Well, of course I don't want to deny these things. But this doesn't necessarily mark out visual experiences or pains as different in kind from other physical processes.

It is tempting to think of our introspective gaze as being attracted by some kind of inner illumination. The reason some states, but not others, are accessible, we suppose, is because they glow with a special light. But this is not the only way to see things. By way of analogy, consider the items that appear on the television news. We don't think that they are distinguished from the ordinary run of events by some distinctive radiance. They are just events that happen to attract the attention of the cameras. Similarly there is no reason to think of our conscious states as being distinguished by some extra lustre. They appear to us as they do in virtue of our having access to them, not because they have some distinctive luminosity.

In some ways, the picture I am here recommending is not unlike panpsychism. In my view, the idea of phenomenally 'being like something', as opposed to being introspectively accessible, fails to draw a line in nature. We should not think of this idea as distinguishing events lit up by phenomenology from those that are mere darkness. But I adopt this even-handedness, not because I think something needs to be added to the physical realm, as do orthodox panpsychists, but because I think that physical states as such are already adequate to account for the nature of conscious experience. Ordinary physical states are perfectly well-

qualified to be like something for subjects. To achieve this, they need only play a role in reasoning and action planning, and so feature in the integrated mental lives of the beings that have them.

## References

---

Alter, T. and Coleman, C. (2018), 'Panpsychism and Russellian Monism' in W. Seager (ed.), *Routledge Handbook of Panpsychism*. London: Routledge.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Armstrong, D. (1968), *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Ball, D. (2009), 'There Are No Phenomenal Concepts', *Mind* 118: 935–62.

[Google Scholar](#) [WorldCat](#)

Bird, A. (2007), *Nature's Metaphysics. Laws and Properties*. Oxford: Clarendon Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Block, N. (1978), 'Reductionism: Philosophical Analysis' in *Encyclopedia of Bioethics*. London: Macmillan.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Block, N. (1995), 'On a Confusion about a Function of Consciousness', *Behavioral and Brain Sciences* 18: 227–87.

[Google Scholar](#) [WorldCat](#)

Block, N. and Stalnaker, R. (1999), 'Conceptual Analysis, Dualism, and the Explanatory Gap', *Philosophical Review* 108: 1–46.

Chalmers, D. (2002), 'Does Conceivability Entail Possibility?' in J. Hawthorne and T. Gendler (eds), *Conceivability and Possibility*. Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Chalmers, D. and Jackson, F. (2001), 'Conceptual Analysis and Reductive Explanation', *Philosophical Review* 110: 315–61.

[Google Scholar](#) [WorldCat](#)

Davidson, D. (1970), 'Mental Events' in L. Foster and J. Swanson (eds), *Experience and Theory*. London: Duckworth.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Dennett, D. (1992), *Consciousness Explained*. London: Allen Lane.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

p. 35 Feigl, H. (1958), 'The "Mental" and the "Physical"' in H. Feigl, M. Scriven, and G. Maxwell (eds), *Minnesota Studies in the Philosophy of Science vol II*. Minneapolis: University of Minnesota Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Goff, P. (2011), 'A Posteriori Physicalists Get Our Phenomenal Concepts Wrong', *Australasian Journal of Philosophy* 89: 191–209.

[Google Scholar](#) [WorldCat](#)

Goff, P. (2017), *Consciousness and Fundamental Reality*. Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Howell, R. (2015). 'The Russellian Monist's Problems with Mental Causation', *Philosophical Quarterly* 65: 22–39.

[Google Scholar](#) [WorldCat](#)

Jackson, F. (1986), 'What Mary Didn't Know', *Journal of Philosophy* 83: 291–5.

[Google Scholar](#) [WorldCat](#)

Jackson, F. (2007), 'The Knowledge Argument, Diaphonousness, Representationalism' in T. Alter and S. Walter (eds), *Phenomenal*

*Concepts and Phenomenal Knowledge*. Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Lee, G. (2014), 'Materialism and the Epistemic Significance of Consciousness' in U. Kriegel (ed.), *Current Controversies in Philosophy of Mind*. London: Routledge.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Levine, J. (1983), 'Materialism and Qualia: The Explanatory Gap', *Pacific Philosophical Quarterly* 64: 354–61.

[Google Scholar](#) [WorldCat](#)

Levine, J. (2001), *Purple Haze. The Puzzle of Consciousness*. Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Levine, J. (2010), 'The Q Factor: Modal Rationalism vs Modal Autonomism', *Philosophical Review* 119: 365–80.

[Google Scholar](#) [WorldCat](#)

Lewis, D. (1966), 'An Argument for the Identity Theory', *Journal of Philosophy* 63: 17–25.

[Google Scholar](#) [WorldCat](#)

Lewis, D. (1988), 'What Experience Teaches', *Proceedings of the Russellian Society of Sydney University* 13: 29–57.

[Google Scholar](#) [WorldCat](#)

Lockwood, M. (1993), 'The Grain Problem' in H. Robinson (ed.), *Objections to Physicalism*. Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

McGinn, C. (1991), *The Problem of Consciousness*. Oxford: Basil Blackwell.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Menzies, P. (2008), 'Causal Exclusion, the Determination Relation, and Contrastive Causation' in J. Kallestrup and J. Hohwy (eds), *Being Reduced: New Essays on Reductive Explanation and Special Science Causation*. Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Milner, A. and Goodale, M. (2008), 'Two Visual Systems Reviewed', *Neuropsychologia* 46: 774–85.

[Google Scholar](#) [WorldCat](#)

Nagel, T. (1986), *The View from Nowhere*. Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Nida-Rümelin, N. (2007), 'Grasping Phenomenal Properties' in T. Alter and S. Walter (eds), *Phenomenal Concepts and Phenomenal Knowledge*, Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Oppenheim, P. and Putnam, H. (1958), 'Unity of Science as a Working Hypothesis' in H. Feigl, M. Scriven, and G. Maxwell (eds), *Minnesota Studies in the Philosophy of Science vol II*. Minneapolis: University of Minnesota Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Papineau, D. (1993), 'Physicalism, Consciousness, and the Antipathetic Fallacy', *Australasian Journal of Philosophy* 71:169–83.

[Google Scholar](#) [WorldCat](#)

Papineau, D. (2002), *Thinking about Consciousness*. Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Papineau, D. (2007), 'Kripke's Argument is Ad Hominem Not Two-Dimensional' in *Philosophical Perspectives* 21: 475–94.

[Google Scholar](#) [WorldCat](#)

Papineau, D. (2010), 'What Exactly is the Explanatory Gap?', *Philosophia* 39: 5–19.

[Google Scholar](#) [WorldCat](#)

Place, U. (1956), 'Is Consciousness a Brain Process?', *British Journal of Psychology* 47: 44–50.

[Google Scholar](#) [WorldCat](#)

Robinson, W. (2015), 'Epiphenomenalism' *Stanford Encyclopedia of Philosophy*.

[Google Scholar](#) [WorldCat](#)

Russell, B. (1927), *The Analysis of Matter*. London: Kegan Paul.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Shea, N. and Bayne, T. (2010), 'The Vegetative State and the Science of Consciousness', *British Journal for the Philosophy of Science* 61: 459–84.

[Google Scholar](#) [WorldCat](#)

p. 36 Smart, J. (1959), 'Sensations and Brain Processes', *Philosophical Review* 68: 141–56.

[Google Scholar](#) [WorldCat](#)

Stoljar, D. (2006), *Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness*. Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Strawson, G. (2003), 'Real Materialism' in G. Strawson (ed.), *Real Materialism and Other Essays*. Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Tye, M. (2009), *Consciousness Revisited*. Cambridge, MA: MIT Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Woodward, J. (2005), *Making Things Happen*. Oxford: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

## Notes

- 1 Perhaps the thesis of 'causal' closure is better formulated as the claim that (the chances of) all physical events are *determined* by prior physical events according to physical law. On some views of causation, these prior determiners do not necessarily count as causes (Woodward 2005; Menzies 2008). Still, determinational closure itself sustains an argument against metaphysically independent mental influences. I shall ignore this complication in what follows.
- 2 In particular, epiphenomenalism is arguably *self-stultifying*, in that it implies that the conscious realm has no casual impact on the views of those who believe in it (Robinson 2015: sect 2.4).
- 3 For a more detailed account of the history of causal arguments for physicalism since the seventeenth century, see the Appendix to my *Thinking about Consciousness* (2002).
- 4 See Feigl 1958; Oppenheim and Putnam 1958; Smart 1959; Lewis 1966; Armstrong 1968; Davidson 1970.
- 5 Among the exceptions who resist phenomenal concepts are David Lewis (1988) and Daniel Dennett (1992). Curiously, they have now been joined by Jackson himself, who has come to view dualism as untenable while continuing to maintain that it would follow from the existence of phenomenal concepts (2007). Other recent writers reject phenomenal concepts on the grounds that no concepts can constitutively depend on prior conscious experiences (Ball 2009; Tye 2009); in my view, this argument sets the standards for phenomenal concepts too high.
- 6 In any case, a derivation from the strictly *physical* facts alone was never really on the cards, given the presence of such observational terms as *white* in the conceptual role of *salt*. See Levine 2010.
- 7 For this line of objection see Howell (2015), and for a Russellian response see Alter and Coleman (2018).
- 8 This is not to deny that some coarse-grained theoretical roles—that of an electrical insulator, say—can be variably realized by different states of affairs with different fine-grained specifications. But that isn't enough for the Russellian monist, who



needs even the most fine-grained theoretical roles to be variably realized.

- 9 Doesn't a persistent intuition of dualism itself amount to an argument against physicalism? No. E is only evidence against T if E is less probable given T than not-T. However, if the alternatives to physicalism are epiphenomenalism or Russellian monism, then they make dualist intuitions no more likely than physicalism does, at least insofar as we think of these intuitions as publicly expressed. After all, it is agreed on all sides that the conscious realm makes no causal difference to views about it: epiphenomenalists and Russellian monists agree with physicalists about the way the brain works, and so in particular about the processes that give rise to expressions of intuitive dualism. (Cf footnote 2 above.)
- 10 For a more detailed discussion of the literature on explanations for dualist intuitions, see Papineau (2010 sect 7). I would like to thank Dara Ghaznavi for drawing my attention to Place's 'phenomenological fallacy'.
- 11 Lee 2014 argues similarly that a committed physicalism undermines any distinctive role for phenomenal consciousness. But see Shea and Bayne (2010) for an attempt to identify such a role.