

# Physicalism

*First published Tue Feb 13, 2001; substantive revision Tue May 25, 2021*

Physicalism is, in slogan form, the thesis that everything is physical. The thesis is usually intended as a metaphysical thesis, parallel to the thesis attributed to the ancient Greek philosopher Thales, that everything is water, or the idealism of the 18th Century philosopher Berkeley, that everything is mental. The general idea is that the nature of the actual world (i.e. the universe and everything in it) conforms to a certain condition, the condition of being physical. Of course, physicalists don't deny that the world might contain many items that at first glance don't seem physical — items of a biological, or psychological, or moral, or social, or mathematical nature. But they insist nevertheless that at the end of the day such items are physical, or at least bear an important relation to the physical.

- [1. Preliminaries](#)
  - [1.1 Terminology](#)
  - [1.2 Historical Issues](#)
  - [1.3 A Framework for Discussion](#)
- [2. The Completeness Question](#)
  - [2.1 Supervenience and Necessity Physicalism](#)
  - [2.2 Identity Physicalism](#)
    - [2.2.1 Token Physicalism](#)
    - [2.2.1 Type Physicalism](#)
  - [2.3 Realization Physicalism](#)
    - [2.3.1 Second-order Physicalism](#)
    - [2.3.2 Subset Physicalism](#)
  - [2.4 Grounding Physicalism](#)
  - [2.5 Fundamentality Physicalism](#)
- [3. Varieties of Physicalism](#)
  - [3.1 Reductive and Non-Reductive Physicalism](#)
  - [3.2 A Priori and A Posteriori Physicalism](#)
- [4. The Condition Question](#)
  - [4.1 The Theory and Object Conceptions of the Physical](#)
  - [4.2 Circularity](#)
  - [4.3 Hempel's dilemma](#)
  - [4.4 The Panpsychism Problem](#)
  - [4.5 The Via Negativa](#)
  - [4.6 Structuralist Approaches](#)
  - [4.7 Physicalism as an Attitude](#)
- [5. The Case Against Physicalism](#)
  - [5.1 Qualia and Consciousness](#)
  - [5.2 Meaning and Intentionality](#)
  - [5.3 Numbers and Abstracta](#)
  - [5.4 Methodological Issues](#)
- [6. The Case for Physicalism](#)
- [Bibliography](#)
- [Academic Tools](#)
- [Other Internet Resources](#)
- [Related Entries](#)

## 1. Preliminaries

### 1.1 Terminology

Physicalism is sometimes known as 'materialism'. Indeed, on one strand to contemporary usage, the terms 'physicalism' and 'materialism' are interchangeable. But the two terms have very different histories. The word 'materialism' appears in English towards the end of the 17th century, but the word 'physicalism' was introduced into philosophy only in the 1930s by [Otto Neurath](#) (1931) and [Rudolf Carnap](#) (1959/1932), both of whom were key members of the Vienna Circle, a group of philosophers, scientists and mathematicians active in Vienna prior to World War II. While it is not clear that Neurath and Carnap understood physicalism in the same way, one thesis often attributed to them (e.g. in Hempel 1949) is the linguistic thesis that every statement is synonymous with (i.e. is equivalent in meaning with) some physical statement. But materialism as traditionally construed is not a linguistic thesis at all; rather it is a metaphysical thesis in the sense that it tells us about the nature of the world. At least for the positivists, therefore, there was a clear reason for distinguishing physicalism (a linguistic thesis) from materialism (a metaphysical thesis). Moreover, this reason was compounded by the fact that, according to official positivist doctrine, metaphysics is nonsense. Since the 1930s, however, the positivist philosophy that under-girded this distinction has for the most part been rejected—for example, physicalism is not a linguistic thesis for

contemporary philosophers—and this is one reason why the words ‘materialism’ and ‘physicalism’ are now often interpreted as interchangeable.

Some philosophers suggest that ‘physicalism’ is distinct from ‘materialism’ for a reason quite unrelated to the one emphasized by Neurath and Carnap. As the name suggests, materialists historically held that everything was matter — where matter was conceived as “an inert, senseless substance, in which extension, figure, and motion do actually subsist” (Berkeley, *Principles of Human Knowledge*, par. 9). But physics itself has shown that not everything is matter in this sense; for example, forces such as gravity are physical but it is not clear that they are material in the traditional sense (Lange 1865, Dijksterhuis 1961, Yolton 1983). So it is tempting to use ‘physicalism’ to distance oneself from what seems a historically important but no longer scientifically relevant thesis of materialism, and related to this, to emphasize a connection to physics and the physical sciences. However, while physicalism is certainly unusual among metaphysical doctrines in being associated with a commitment both to the sciences and to a particular branch of science, namely physics, it is not clear that this is a good reason for calling it ‘physicalism’ rather than ‘materialism.’ For one thing, many contemporary physicalists do in fact use the word ‘materialism’ to describe their doctrine (e.g. Smart 1963). Moreover, while ‘physicalism’ is no doubt related to ‘physics’ it is also related to ‘physical object’ and this in turn is very closely connected with ‘material object’, and via that, with ‘matter.’

In this entry, I will adopt the policy of using both terms interchangeably, though I will typically refer to the thesis we will discuss as ‘physicalism’. It is important to note, though, that physicalism (aka materialism) is often associated with a number of further doctrines distinct from the one will focus on; for a discussion of some of these, see the supplement on:

[Physicalism and Associated Doctrines](#)

## 1.2 Historical Issues

Setting aside what it is properly called, the thesis of physicalism is often described as an extremely old, even ancient, thesis. The first sentence of [Friedrich Lange](#)’s *The History of Materialism*, which was the standard work on the subject in the 19th century is: “Materialism is as old as philosophy, but not older” (1925, 3). What Lange has in mind is the pre-Socratic philosopher [Democritus](#), who is usually thought of defending a kind of physicalism or materialism when he said or allegedly said, “all is atoms and the void.” This view casts a long shadow over subsequent formulations of physicalism. A position like that of Democritus was revived in the early modern period just prior to Newton, by philosophers and scientists such as [Hobbes](#) or [Gassendi](#). In the eighteenth century, French philosophers like [D’Holbach](#) and La Mettrie thought of themselves as materialists (and would be now classified as physicalists) somewhat in the same way since they held that each human being is a complicated sort of machine. In the nineteenth century, while [Karl Marx](#)’s ‘dialectical materialism’ is something different from the metaphysical thesis we will focus on, he nevertheless developed his social philosophy against the background of what we would call physicalism; in fact, Marx’s doctoral dissertation was a comparison of Democritus and [Epicurus](#). And in the twentieth century, analytic philosophers such as Smart and Lewis self-consciously defended their views in a way that acknowledges, as Lewis put it, “our intellectual ancestors” (1994, 293).

However, while there is certainly something in common here, the underlying historical issues are extremely complicated, since they involve subtle questions of scientific and philosophical change. We won’t discuss them in detail in this entry. It is worth emphasis, though, that we should be careful in lumping different people in different epochs together. The precise views they held are often different from one another. As we noted above, and as Lange himself emphasises, Newton did not think that, even in the physical world, all is atoms and the void, since for him there are also forces such as gravity. It follows that any post-Newtonian philosophers who think of themselves as physicalists must have a different thesis in mind from Democritus.

## 1.3 A Framework for Discussion

In approaching the topic of physicalism, one may distinguish what I will call *the interpretation question* from *the truth question*. The interpretation question asks:

- What does it *mean* to say that everything is physical?

The truth question asks:

- Is it *true* to say that everything is physical?

There is obviously a sense in which the second question presupposes an answer to the first — you need to know what a statement means before you can ask whether it’s true — and so we will begin with the interpretation question.

The interpretation question itself divides into two sub-questions, which I will call *the condition question* and *the completeness question*. The condition question asks:

- What does it mean to for something to be *physical*?

The condition question holds fixed the issue of what it means for everything to satisfy some condition, or to bear a relation to something that satisfies that condition, and asks instead what *is* the condition, *being physical*, that everything satisfies or bears a relation to. Notice that a parallel question could be asked of Thales: what is the condition, *being water*, that according to Thales, everything satisfies?

The completeness question asks:

- What relation or relations must obtain between *everything* and the physical if physicalism is to be true?

In other words, the completeness question holds fixed the issue of what it means for something to be physical, and asks instead what relation or relations obtain between everything and the physical if physicalism is true; in what sense, in other words, is physicalism a complete thesis, a thesis

that applies to everything whatsoever. Notice again that a parallel question could be asked of Thales: assuming we know what condition you have to satisfy to be water, what does it mean to say that everything satisfies that condition?

Once again there is a sense in which the second question here presupposes an answer to the first — you need to know what it is for something to be physical in order to assess different proposals about the relation everything bears to the physical. Nevertheless, it is easier from a presentational point of view to discuss the completeness question first, leaving our answer to the condition question for the moment impressionistic, and that will be our procedure.

## 2. The Completeness Question

How should we approach the completeness question? In the history of attempts to answer this question, people have tended to adopt one of two strategies. One appeals to (what philosophers call) *modal* notions, where a modal notion here means a notion connected to possibility and necessity, to what might or must be the case. Another appeals to *non-modal* notions, that is, notions distinct from ideas about possibility and necessity. There is a wide variety of such notions, though perhaps the most obvious one is *identity* in the logical sense, according to which if  $x$  is identical to  $y$ , then every property of  $x$  is a property of  $y$ . This notion has consequences for possibility and necessity, but it is not itself modal. Other non-modal notions include *realization*, or *grounding*, which we will consider below.

In practice, most formulations of physicalism include *both* modal and non-modal elements at some level. Proponents of modal formulations often end up appealing to something non-modal in the course of elaborating and defending their view. Non-modal versions are usually interpreted so that they have the same modal consequences as modal versions. Nevertheless, the issue of whether to answer the completeness question by appealing to modal or non-modal ideas has proved as controversial as any in the literature on what physicalism is (for overviews, see Rabin 2020, Elpidorou 2018a, Tiehen 2018). In what follows, we will first look at an influential modal answer to the completeness question, which appeals to *supervenience*, and then turn to several non-modal alternatives.

### 2.1 Supervenience and Necessity Physicalism

The idea of supervenience has its origins in meta-ethics but was imported into philosophy of mind mainly by Davidson 1970; for a survey, see [supervenience](#). For our purposes, the general idea might be introduced via an example due to David Lewis of a dot-matrix picture:

A dot-matrix picture has global properties — it is symmetrical, it is cluttered, and whatnot — and yet all there is to the picture is dots and non-dots at each point of the matrix. The global properties are nothing but patterns in the dots. They supervene: no two pictures could differ in their global properties without differing, somewhere, in whether there is or there isn't a dot (1986, p. 14).

This gives us one way to think about the basic idea of physicalism. The basic idea is that the physical features of the world are like the dots in the picture, and the psychological or biological or social features of the world are like the global properties of the picture. Just as the global features of the picture supervene on the dots, so too everything supervenes on the physical, if physicalism is true.

It is desirable to have a more explicit statement of physicalism, and here too Lewis's example gives us direction. He says that, in the case of the picture, supervenience means that “no two pictures can be identical in the arrangement of dots but different in their global properties”. Similarly, one might say that, in the case of physicalism, no two possible worlds can be identical in their physical properties but differ, somewhere, in their mental, social or biological properties. To weaken this slightly, we might say that if physicalism is the case at *our* world — that is, is true of our universe and everything in it — then no *other* world can be physically identical to our world without being identical to it in all respects. This suggests the following general account of what physicalism is (in the following formulation and in subsequent ones, we use “iff” to abbreviate “if and only if”):

- (1) Physicalism is true at a possible world  $w$  iff any world which is a physical duplicate of  $w$  is a duplicate of  $w$  *simpliciter*.

If physicalism is construed along the lines suggested in (1), we have an answer to the completeness question. The completeness question asks: what relation does everything bear to the physical if physicalism is true? According to (1), the answer is that everything must supervene on the physical; or, to put it more technically, there is no possible world which is identical to our world in every physical respect but which is not identical to it in a biological or social or psychological respect. It will be useful to have a name for physicalism so defined, so let us call it *supervenience physicalism*.

Supervenience offers one modal formulation of physicalism, but it is worth taking note of a second modal formulation too. Suppose we say that a property  $G$  is necessitated by a property  $F$  just in case, in all possible worlds, if something is  $F$  then it is  $G$ ; in this sense, for example, being red necessitates being colored, and being square necessitates having some extension in space. This suggests a formulation of physicalism along the following lines:

- (2) Physicalism is true at a possible world  $w$  iff every property instantiated at  $w$  is necessitated by a physical property.

What is the relation between physicalism so defined, which we might call *necessity physicalism*, and *supervenience physicalism*? At least if necessitation is understood as a sort of entailment, then these are not equivalent; for discussion of this point, see [supervenience](#). However, (1) and (2) are clearly similar, in particular they are modal formulations of physicalism. In what follows we will concentrate on supervenience physicalism, but what we will say will apply also to necessity physicalism.

Supervenience physicalism was for many years the dominant version of physicalism; perhaps because of this, many different problems have been raised for it; some of these problems are discussed in the supplement on:

[Supervenience Physicalism: Further Issues](#)

But the most influential objection to supervenience physicalism (and to modal formulations generally) is what might be called *the sufficiency problem*. This alleges that, while (1) articulates a necessary condition for physicalism it does not provide a sufficient condition. The underlying

rationale is that, intuitively one thing can supervene on another and yet be of a completely different nature. To use Fine's famous (1994) example, consider the difference between Socrates and his singleton set, the set that contains only Socrates as a member. The facts about the set supervene on the facts about Socrates; any world that is like ours in respect of the existence of Socrates is like ours in respect of the existence of his singleton set. And yet the set is quite different from Socrates. This in turn raises the possibility that something might be of a completely different nature from the physical and nevertheless supervene on it.

One may bring out this objection further by considering positions in philosophy which entail supervenience and yet deny physicalism. A good example is *necessitation dualism*, which is an approach that weaves together elements of both physicalism and its traditional rival, dualism. On the one hand, the necessitation dualist wants to say that mental facts and physical facts are metaphysically distinct—just as a standard dualist does. On the other hand, the necessitation dualist wants to agree with the physicalist that mental facts are necessitated by, and supervene on, the physical facts. If this sort of position is coherent, (1) does not articulate a sufficient condition for physicalism. For if necessitation dualism is true, any physical duplicate of the actual world is a duplicate *simpliciter*. And yet, if dualism of any sort is true, including necessitation dualism, physicalism is false.

How to respond to the sufficiency problem? Some respond by denying the coherence of the position that causes the problem. Necessitation dualism as we have just described it violates (what is known as) Hume's dictum that there are no necessary connections between distinct existences. According to necessitation dualism, mental and physical properties are metaphysically distinct, and yet are necessarily connected. However, Hume's dictum is itself a matter of controversy (see Jackson 1993, Stalnaker 1996, Stoljar 2010, and Wilson 2005, 2010). Another approach appeals to a priori physicalism which will examine below (see Jackson 2006). But by far the most common response has been to concede that the sufficiency problem shows that supervenience formulation of physicalism is too weak (e.g. Kim 1998), and to look for an alternative.

## 2.2 Identity Physicalism

Suppose then that (1) provides a necessary condition for physicalism but not a sufficient condition; how might we strengthen it to make it more plausible? The most obvious thing to do is to appeal to identity. Indeed, in the history of attempts to answer the completeness question, the appeal to identity predates the appeal to supervenience. Nevertheless, this version of physicalism—*identity physicalism* as we may call it—runs into serious problems.

### 2.2.1. Token Physicalism

In fact there are two different versions of identity physicalism, *type physicalism* and *token physicalism*. Token physicalism is the view that every particular thing in the world is a physical particular. So the token physicalist says that physicalism should be formulated in the following way:

- (3) Physicalism is true at a possible world  $w$  iff for every particular (object, event or process)  $x$  that exists at  $w$ , there is some physical particular  $y$  such that  $x = y$

But (3) offers neither necessary nor sufficient conditions for physicalism. To see that it is not sufficient, consider the variety of dualism usually called *property dualism*. Property dualism says that (a) every particular is a physical particular but (b) some particulars (e.g. human beings) have psychological properties wholly distinct from any physical properties. The contrast here is with *substance dualism*. The substance dualist agrees with the property dualist that some particulars have psychological properties wholly distinct from any physical properties, but they will add that such particulars are themselves non-physical.

Token physicalism – physicalism according to (3) – is certainly inconsistent with substance dualism. Substance dualism entails that some particulars are non-physical, token physicalism denies it. But token physicalism is compatible with property dualism; indeed property dualism entails that token physicalism is true. On the other hand, property dualism is usually understood as being inconsistent with physicalism in any form. Hence token physicalism is not sufficient for physicalism.

The problem that property dualism presents for token physicalism is noteworthy in several respects. For one thing, it is similar to the problem necessitarian dualism presents for supervenience physicalism, though necessitarian dualism can itself be developed either as a sort of property dualism or as a sort of substance dualism. For another thing, it brings out the important role properties rather than particulars play in the contemporary discussion of physicalism. If one ignores properties, the dispute between physicalism and dualism may easily be understood as the dispute between token physicalism and substance dualism; but once properties are factored in, things look very different.

Not only does (3) not provide a sufficient condition for physicalism, it does not provide a necessary condition either. Consider a social or legal object such as the United States Court of Appeals for the Seventh Circuit. According to (3), if physicalism is true, there must be some physical object or particular for the court to be identical with. But intuitively, there is no such physical object. Nevertheless, physicalism might still be true. If so, token physicalism is not necessary for physicalism. (Notice here that there is no parallel problem for supervenience physicalism. It entails that the facts about court supervene on physical facts but not that there is any physical object that the court is identical with. For the classic presentation of this point, see Haugeland 1983).

### 2.2.2. Type Physicalism

Turning now to type physicalism, this holds that every property (or at least every property instantiated in the actual world) is identical with some physical property. So the type physicalist supposes that physicalism should be formulated in the following way:

- (4) Physicalism is true at a possible world  $w$  iff every property instantiated at  $w$  is identical to a physical property.

Unlike both token physicalism and supervenience physicalism, type physicalism *is* sufficient for physicalism: if every property instantiated in the actual world is identical with some physical property, then dualism will be false in any of the versions we have considered

However, while (4) provides a sufficient condition for physicalism, it does not provide a necessary condition. Consider again the United States Court of Appeals for the Seventh Circuit. If type physicalism is true, then every property that the court has (for example, having a legal power over lower

courts) must be identical with some physical property. But on the face of it, that is unlikely. Nevertheless, physicalism might still be true. If so, type physicalism is not necessary for physicalism. (Notice again that there is no parallel problem for supervenience physicalism. It entails that the property having a legal power over lower courts supervenes on physical properties, not that it is identical with a physical property.)

Another way to bring out the sense in which type physicalism is too strong for physicalism is to focus on the possibility of *multiple realizability*, as philosophers call it. This is roughly the idea that physically very different creatures can nevertheless share psychological properties; see [multiple realizability](#). A major benefit of supervenience physicalism is that, as usually understood, it is consistent with this possibility. But type physicalism is often thought of as not consistent with it, at least if we focus on instantiated psychological properties. If so, type physicalism may be false while supervenience physicalism may be true.

We have been assuming that supervenience physicalism is distinct from type physicalism. But it is worth noting that, there are ways to understand supervenience according to which this difference is less obvious. The reason for this has to do with questions concerning the logical (or Boolean) closure of the set of physical properties — if  $P$ ,  $Q$  and  $R$  are physical properties, which of the various logical permutations of  $P$ ,  $Q$  and  $R$  are likewise physical properties? On some assumptions concerning closure and supervenience, supervenience physicalism (construed as a necessary truth) entails type physicalism; on other assumptions, it doesn't. But the problem is that the assumptions themselves are difficult to interpret and evaluate, and so the issue remains a difficult one. It is not necessary for our purposes to settle the question concerning closure here. (For further discussion of these issues see Kim 1993, Bacon 1990, Van Cleve 1990, Stalnaker 1996.)

## 2.3 Realization Physicalism

Our discussion of the completeness question has so far yielded negative results: supervenience physicalism is too weak, type physicalism is too strong, and token physicalism is both too weak and too strong. What seems to be required is an approach that, like the modal formulations we started with, entails the supervenience of everything on the physical, but unlike them avoids the sufficiency problem. One prominent idea along these lines appeals to a relation between properties distinct from both identity and supervenience, usually called *realization*. As in the case of identity physicalism, there are two different versions of *realization physicalism* as we may call it; we will consider them in turn.

### 2.3.1 Second-order Physicalism

The first realization definition has been explored and defended in most detail by Andrew Melnyk (see Melnyk 2003 and the references therein). For Melnyk, a property  $F$  realizes a property  $G$  if and only if (a)  $G$  is identical to a second-order property, the property of having some property that has a certain causal or theoretical role; and (b)  $F$  is the property that plays the causal or theoretical role in question. We may call this notion 'second-order realization' to distinguish it from a different notion of realization to be considered in a moment. This suggests:

- (5) Physicalism is true at a possible world  $w$  iff every property instantiated at  $w$  is either a physical property or is second-order realized in a physical property.

Suppose we call physicalism so defined *second-order realization physicalism* or *second-order physicalism* for short; what is the relation between it and supervenience physicalism? Supervenience physicalism does not entail second-order physicalism since the fact that a property  $F$  supervenes on a property  $G$  does not entail that  $F$  is a second-order property.

Does second-order physicalism entail supervenience physicalism? The usual assumption is that it does, but, as Melnyk himself notes at one point (2003, p. 23), there is an issue here having to do with the definition of a second-order property, the property of having some property that has a certain causal or theoretical role. What are the properties involved in spelling out these causal or theoretical roles? If physicalism is true at all, it must be true of these properties as much as any other properties. But then by second-order physicalism, these properties themselves will be either physical or realized by physical properties. If the first option is taken, the second-order physicalist will stand revealed as holding a version of identity physicalism (one level up, as it were), and thus will face the multiple realization objection. If the second option is taken, the second-order physicalist looks committed to an infinite regress, since now we have further properties realized by physical properties and, correlatively, further causal or theoretical roles. To avoid the regress, the second-order physicalist might say that these properties supervene on or are identical to physical properties. But now it's hard to see the difference between the realization physicalist and these other doctrines. (For some further discussion of this issue, see Elpidorou 2018a.)

### 2.3.2 Subset Physicalism

The second realization definition of physicalism has been developed by Wilson 1999, 2011 and Shoemaker 2007. On this view, a property  $F$  realizes a property  $G$  if and only if (a)  $G$  has some set of causal powers or features  $S$ ; (b)  $F$  has some set of causal powers or features  $S^*$ ; and (c)  $S$  is a subset of  $S^*$ . (We may call this notion 'subset realization' to distinguish it from the different notion of realization just considered.) This suggests:

- (6) Physicalism is true at a possible world  $w$  iff every property instantiated at  $w$  is either a physical property or is subset realized in a physical property.

Suppose we call physicalism so defined *subset realization physicalism* or *subset physicalism* for short; what is the relation between it and supervenience physicalism? Supervenience physicalism does not entail subset physicalism since the fact that a property  $G$  supervenes on a property  $F$  does not entail anything about their causal powers. For example, if causation is a macro-phenomenon as some philosophers have held it to be, it may be that  $F$  has no causal powers at all, while  $G$  does.

Does subset realization physicalism entail supervenience physicalism? Well, there is a problem here too having to do with (what is sometimes called) a *causal* theory of properties, that is, a theory according to which the causal powers or features that a property bestows on the things that have it are exhaustive of the nature of that property. Suppose that a causal theory is false. Then, in principle, one property might subset realize another and yet be quite different from it in nature. And this in turn suggests that subset physicalism does not by itself entail supervenience physicalism. Of course, one might respond by asserting that the causal theory is true. But to do that is controversial; indeed, even those philosophers who hold both a subset model *and* a causal theory want to separate out these two commitments (e.g. Shoemaker 2007; see also Wilson 2011). Alternatively, one might

respond by denying that physicalism entails supervenience in the first place, by saying that “lack of ... supervenience is compatible with physicalism” (Wilson 2014, 255, see also Wilson 2011). But this too is controversial; as we saw above, most philosophers assume that supervenience is necessary for physicalism. Hence the status of the subset approach remains controversial.

## 2.4 Grounding Physicalism

An influential recent approach to the completeness question that is different from any we have considered so far focuses on the idea of grounding, something that has been extensively discussed recently in the metaphysics literature (see, e.g., Fine 2001, Schaffer 2009, Rosen 2010, Wilson 2014, Bennett 2017 and the essays in Correia and Schneider 2012). Intuitively, a property  $F$  is grounded in a property  $G$  just in case  $F$  holds in virtue of  $G$ , or the instantiation of  $G$  explains the instantiation of  $F$ . This suggests:

- (7) Physicalism is true at a possible world  $w$  iff every property instantiated at  $w$  is either a physical property or is grounded in a physical property.

Suppose we call physicalism so defined *grounding physicalism*; what is the relation between it and supervenience physicalism? Supervenience physicalism does not entail grounding physicalism, since the fact that a property  $F$  supervenes on a property  $G$  does not entail that  $F$  is grounded by  $G$ .

Does grounding physicalism entail supervenience physicalism? Some philosophers suppose it does (e.g. Rosen 2010) and so for them grounding physicalism would entail supervenience physicalism. But others suppose it does not (e.g. Schaffer 2009) which raises the question of whether a thesis such as (7) by itself provides an account of physicalism, or whether some compromise between it and (1) would have to be reached.

Even if grounding physicalism entails supervenience physicalism, there are further issues about it that have been raised. One problem concerns abstract objects, i.e., entities apparently not located in space and time, such as numbers, properties and relations, or propositions. However, since this problem seems a general problem for all kinds of physicalism, we will discuss it below.

A second problem for grounding physicalism is that the notion of grounding is itself controversial in some quarters. Wilson (2014), for example, points out that grounding per se is similar to supervenience in that it leaves open many of the questions philosophers of mind are interested in, viz., whether the mental exists, whether it is reduced to the physical, and whether it is causally efficacious. She concludes that grounding “cannot do the work” that its proponents want it to do (2014, 542). One might respond that this depends on what work grounding physicalism is supposed to do; indeed, it may be a feature rather than a bug that grounding leaves these things open. But whatever is the truth about this, there is no doubt that the precise contours of the grounding relation are yet to be made out. Hence, the proper assessment of grounding physicalism is at this point a bit unclear (for some further discussion, see, e.g., Berker 2018 and Schaffer 2016).

## 2.5 Fundamentality Physicalism

The final answer to the completeness question we will consider focuses on the idea of a fundamental property, which is a notion discussed extensively in David Lewis’s metaphysics, and the literature that follows on from it. On Lewis’s view, a fundamental or perfectly natural property is a special kind of property, one that is, as he says, “not at all disjunctive, or determinable, or negative. They render their instances perfectly similar in some respect. They are intrinsic; and all other intrinsic properties supervene on them” (2009, 204). This suggests the following formulation of physicalism:

- (8) Physicalism is true at a possible world  $w$  iff every fundamental property instantiated at  $w$  is a physical property.

Suppose we call physicalism so defined *fundamentality physicalism*; what is the relation between it and supervenience physicalism? Supervenience physicalism does not entail fundamentality physicalism, since the fact that a property  $F$  supervenes on a property  $G$  does not entail that either property is fundamental. Nevertheless, it is reasonable to think that fundamentality physicalism entails supervenience physicalism, especially in the light of Lewis’s comment about supervenience just quoted.

How plausible is (8) as an account of physicalism? One objection concerns the notion of a fundamental property quite generally. For at least some philosophers, Lewis’s ideas about fundamentality are, as he himself puts it, a throwback to medieval metaphysics (Lewis 1983). Another objection is that physicalism on this view seems empirically speculative, since it seems to entail that there is a fundamental level in the world; if so, definitions in terms of grounding may do better (Schaffer 2003). Whatever is the truth about these objections, it is an interesting historical fact that Lewis defines physicalism twice-over. He defines it as supervenience physicalism (as we saw above) and also as fundamentality physicalism. There is no suggestion in his work that these are in any sense in tension (for further discussion, see Stoljar 2015). This underscores the point made earlier, namely, that in practice all versions of physicalism, and hence any answer to the completeness question, will include modal and non-modal elements.

## 3. Varieties of Physicalism

We have been considering various answers to the completeness question, namely, what it means to say that everything is physical. At this point, it is worth considering two issues associated with completeness that have not so far been brought to the surface, and which suggest different varieties of physicalism. The first is whether physicalism involves a kind of reductionism; the second is whether physicalism involves what philosophers call a priori entailment or deducibility.

### 3.1 Reductive and Non-reductive Physicalism

The main problem in assessing whether a physicalist must be a reductionist is that there are various non-equivalent versions of reductionism.

One idea is tied to the notion of conceptual or reductive analysis. When philosophers attempt to provide an analysis of some concept or notion, they often try to provide a reductive analysis of the notion in question, i.e. to analyze it in other terms. Applied to the philosophy of mind, this notion

might be thought of entailing the idea that every mental concept or predicate is analyzed in terms of a physical concept or predicate. A formulation of this idea is (9):

- (9) Reductionism is true iff for each mental predicate  $F$ , there is a physical predicate  $G$  such that a sentence of the form ‘ $x$  is  $F$  iff  $x$  is  $G$ ’ is analytically true.

While one occasionally finds in the literature the suggestion that physicalists are committed to (9) in fact, no physicalist since before Smart (1959) has (unqualifiedly) held anything like it. Adapting Ryle (1949), Smart supposed that in addition to physical expressions there is a class of expressions which are topic-neutral, i.e. expressions which were neither mental nor physical but when conjoined with any theory would greatly increase the expressive power of the theory. Smart suggested that one might analyze mental expressions in topic-neutral (but not physical) terms, which in effect means that a physicalist could reject (9).

A different notion of reduction derives from the attempts of philosophers of science to explain intertheoretic reduction. The classic formulation of this notion was given by Ernest Nagel (1961). Nagel said that one theory was reduced to another if you could logically derive the first from the second together with what he called bridge laws, i.e., laws connecting the predicates of the reduced theory (the theory to be reduced) with the predicates of the reducing theory (the theory to which one is reducing). Here is a formulation of this idea, where the theories in question are psychology and neuroscience:

- (10) Reductionism is true iff for each mental predicate  $F$  there is a neurobiological predicate  $G$  such that a sentence of the form ‘ $x$  is  $F$  iff  $x$  is  $G$ ’ expresses a bridge law.

Once again, however, there is no reason at all why physicalists need to accept that reductionism is true in the sense of (10). We noted earlier that the possibility of multiple realization renders type physicalism implausible; something similar is true in this case. Many different neurobiological processes (whether in our own species or a different one) could underlie the same psychological process — indeed, given science fiction, even non-neurobiological processes might underlie the same psychological process. If so, (10) seems to be false. (For a classic presentation of multiple realization and reductionism, see Fodor 1974, but for an alternative view, see Kim 1993).

A third notion of reductionism is more metaphysical in focus than either the conceptual or theoretical ideas reviewed so far. According to this notion, reductionism means that the properties expressed by the predicates of (say) a psychological theory are identical to the properties expressed by the predicates of (say) a neurobiological theory — in other words, this version of reductionism is in essence a version of type physicalism or the identity theory. As we have already seen, however, since a physicalism need not be a type physicalism, it need not be reductionism in this metaphysical sense either.

A final notion of reductionism that needs to be distinguished from the previous three concerns whether mental statements follow *a priori* from non-mental statements. Here is a statement of this sort of idea,

- (11) Reductionism is true iff for each mental predicate  $F$  there is non-mental predicate  $G$  such that a sentence of the form “if  $x$  is  $F$  then  $x$  is  $G$ ” is *a priori*.

What (11) says is that if reductionism is true, *a priori* knowledge alone, plus knowledge of the physical or non-mental truths will allow one to know the mental truths. This question is usually debated in the context of another, viz., the question of *a posteriori* and *a priori* physicalism. So we now turn to that question.

### 3.2 *A Priori* and *A Posteriori* Physicalism

We have been assuming that supervenience is necessary for physicalism, even if it is not sufficient. But what follows from supervenience? One thing that is usually thought to follow is that the physical truths of the world entail *all* the truths; hence physicalism is true at all (12) is true

- (12) The physical truths entail all the truths.

Now suppose that  $S$  is a statement which specifies the physical nature of the actual world and  $S^*$  is a statement which specifies the total nature of the world. (It might be that neither  $S$  nor  $S^*$  are expressible in languages we can understand, but let us set this aside.) Then physicalism tells us in addition that that:

- (13)  $S$  entails  $S^*$

Another way to say this is to say that if supervenience physicalism is true, the following conditional is necessarily true:

- (14) If  $S$  then  $S^*$

Indeed, this is a general feature of supervenience physicalism: if it is true then there will always be a necessary truth of the form of (14).

Now, if (14) is necessary the question arises whether it is *a priori*, i.e. knowable independent of empirical experience, or whether it is *a posteriori*, i.e. knowable but not independently of empirical experience. Traditionally, every statement that was necessary was assumed to be *a priori*. However, since Kripke’s *Naming and Necessity* (1980), philosophers have become used to the idea that there are truths which are both necessary and *a posteriori*. Accordingly many recent philosophers have defended a *posteriori* physicalism: the claim that statements such as (14) are necessary and *a posteriori* (cf. Loar 1997). Moreover, they have used this point to try to disarm many objections to physicalism, including those concerning qualia and intentionality that we will consider below. Moreover, as we have just noted, some philosophers have suggested that the necessary *a posteriori* provides the proper interpretation of non-reductive physicalism.

The appeal to the necessary *a posteriori* is on the surface an attractive one, but it is also controversial. One problem arises from the fact that Kripke’s



idea that there are necessary and *a posteriori* truths can be interpreted in two rather different ways. On the first interpretation — I will call it the derivation view — while there are necessary *a posteriori* truths, these truths can be derived *a priori* from truths which are *a posteriori* and contingent. On the second interpretation — I will call it the non-derivation view — there are *non-derived* necessary *a posteriori* truths, i.e. necessary truths which are not derived from any contingent truths (or any *a priori* truths for that matter). The problem is that when one combines the derivation view with the claim that (14) is necessary and *a posteriori*, one encounters a contradiction. If the derivation view is correct, then there is some contingent and *a posteriori* statement  $S\#$  that logically entails (14). However, if  $S\#$  logically entails (14) then (since ‘If  $C$ , then if  $A$  then  $B$ ’ is equivalent to ‘If  $C \& A$ , then  $B$ ’) we can infer that the following is both necessary and *a priori*:

(15) If  $S \& S\#$  then  $S^*$ .

One the other hand, if physicalism is true, and  $S$  summarizes the total nature of the world then  $S\#$  is *already implicitly included in  $S$* . In other words, (15) is simply an expansion of (14). But if (15) is just an expansion of (14), then if (15) is *a priori*, (14) must also be *a priori*. But that means our initial assumption is false: (14) is not a necessary *a posteriori* truth after all (see Jackson 1998).

How might an *a posteriori* physicalist respond to this objection? The obvious response is to reject the derivation view of the necessary *a posteriori* in favor of the non-derivation view. But this is just to say that if one wants to defend *a posteriori* physicalism, one will have to defend the non-derivation view of the necessary *a posteriori*. However, the non-derivation view is controversial, so it is not something that we can hope to solve here. (For discussion, see Byrne 1999, Chalmers 1996, 1999, Jackson 1998, Loar 1997, Lewis 1994, Yablo 1999, and the papers in Gendler and Hawthorne 2004)

## 4. The Condition Question

Earlier we distinguished two interpretative questions with respect to physicalism, the completeness question and the condition question. So far we have been concerned with the completeness question. I turn now to the condition question, the question of what it is for something (an object, an event, a process, a property) to be physical.

This issue has received less attention in the literature than the questions we have been studying so far. But it is just as important. Without any understanding of what the physical is, we can have no serious understanding of what physicalism is. After all, what does it mean to say that everything is physical, as opposed to chemical or financial. (The point here is quite general: if Thales says that everything is water, we don’t understand what he says unless he says something about what water is. The physicalist is in the same boat.)

### 4.1 The Theory and Object Conceptions of the Physical

So what is the answer to the condition question? If we concentrate for simplicity on the notion of a physical property, we can discern two kinds of answers to this question (see Stoljar 2001, Stoljar 2010, Tiehen 2018). The first ties the notion of a physical property to a notion of a physical theory, for this reason we can call it the theory-based conception of a physical property:

#### **The theory-based conception:**

A property is physical iff it is the sort of property that physical theory tells us about.

According to the theory-based conception, for example, if physical theory tells us about the property of having mass, then having mass is a physical property. (The theory-based conception bears some relation to the notion of physicalism discussed in Feigl 1967; more explicit defense is found in Smart 1978, Lewis 1994, Braddon-Mitchell and Jackson 1996, and Chalmers 1996.)

The second kind of answer ties the notion of a physical property to the notion of a physical object, for this reason we can call it the object-based conception of a physical property:

#### **The object-based conception:**

A property is physical iff it is the sort of property had by paradigmatic physical objects and their constituents.

According to the object-based conception, for example if paradigmatic physical objects such as rocks or planets are solid or are located in space or have mass, then the property of being solid or located in space or having mass is a physical property. (The best examples of philosophers who operate with the object-conception of the physical are Meehl and Sellars 1956 and Feigl 1967; more recent defense is to be found in Jackson 1998.)

Do these conceptions characterize the same class of properties? There are a number of different possibilities here, but one that has received some attention in the literature is that physical theory only tells us about the dispositional properties of physical objects, and so does not tell us about the categorical properties, if any, that they have — a thesis of this sort has been defended by a number of philosophers, among them Russell (1927), Armstrong (1968), Blackburn (1992) and Chalmers (1996). Hence, if physical objects do have categorical properties, those properties will not count as physical by the standards of the theory conception. On the other hand, there seems no reason not to count them as physical in some sense or other. If that is right, however, then the possibility emerges that the theory- and the object-conceptions characterize distinct classes of properties.

A further distinction that can be made here, and which cross-cuts the theory/object distinction, is between a more narrow and a more broad conception of a physical property (see Chalmers 2015). A narrow conception counts as a physical property only those within a restricted range. On the theory conception, for example, these might be the properties expressed by fundamental physics. A broader conception counts as physical property any property which is either a property in the narrow sense (a narrow physical property, as we might call it) or else bears the right kind of relation to narrow physical properties, for example, is definable in terms of them, or is grounded in them. There are lots of complications here, partly because there are different ways of restricting the narrow class, and different ways of extending such a class. Here we will understand the theory and object conceptions in a narrower rather than a broad way.

### 4.2 Circularity



Along with the concepts of space, time, causality, value, meaning, truth and existence, the concept of the physical is one of the central concepts of human thought. So it should not be surprising that any attempt to come to grips with what a physical property is will be controversial. The theory and object conceptions are no different: each has provoked a number of different questions and criticisms. In the remainder of our discussion of the condition question, I will review some main ones.

To begin with, one might object that both conceptions are inadequate because they are circular, i.e., both appeal to the notion of something physical (a theory or an object) to characterize a physical property. But how can you legitimately explain the notion of one sort of physical thing by appealing to another?

However, the response to this is that circularity is only a problem if the conceptions are interpreted as providing a reductive analysis of the notion of the physical. But there is no reason why they should be interpreted in that way. After all, we have many concepts that we understand without knowing how to analyze (cf. Lewis 1970). So there seems no reason to suppose that either the theory or object conception is providing anything else but a way of understanding the notion of the physical.

The point here is an important one in the context of the condition question. Earlier we said that the condition question was perfectly legitimate because it is legitimate to ask what the condition of being physical is that, according to physicalism, everything has. But this legitimate question should not be interpreted as the demand for a reductive analysis of the notion of the physical. Consider Thales again: it is right to ask Thales what he means by 'water' — and in so doing demand an understanding of the notion of water — but it is wrong to demand of him a conceptual analysis of water.

### 4.3 Hempel's dilemma

One might object that any formulation of physicalism which utilizes the theory-based conception will be either trivial or false. Carl Hempel (cf. Hempel 1969, see also Crane and Mellor 1990) provided a classic formulation of this problem: if physicalism is defined via reference to contemporary physics, then it is false — after all, who thinks that contemporary physics is complete? — but if physicalism is defined via reference to a future or ideal physics, then it is trivial — after all, who can predict what a future physics contains? Perhaps, for example, it contains even mental items. The conclusion of the dilemma is that one has no clear concept of a physical property, or at least no concept that is clear enough to do the job that philosophers of mind want the physical to play.

One response to this objection is to take its first horn, and insist that, at least in certain respects contemporary physics really is complete or else that it is rational to believe that it is (cf. Smart 1978, Lewis 1994 and Melnyk 1997, 2003). There is an element of truth in this. It may be rational to believe that contemporary science is true, even if not that it is complete. Nevertheless, it also seems mistaken to *define* physicalism with respect to the physics that happens to be true in this world. The reason is that whether a physical theory is true or not is a function of the contingent facts; but whether a property is physical or not is not a function of the contingent facts. For example, consider medieval impetus physics. Medieval impetus physics is false (though of course it might not have been) and thus it is irrational to suppose it true. Nevertheless, the property of having impetus — the central property that objects have according to impetus physics — is a physical property, and a counterfactual world completely described by impetus physics would be a world in which physicalism is true. But it is hard to see how any of this could be right if physicalism were defined by reference to the physics that we have now or by the physics that happens to be true in our world. (For development of this point, and for a dilemma that is similar to Hempel's but which casts the issue in modal rather than temporal terms, see Stoljar 2010; for discussion, see Baltimore 2013, Fiorese 2016)

A different response to Hempel's dilemma is that what it shows, if it shows anything, is that a particular proposal about how to define a physical property — namely, via reference to physics at a particular stage of its development — is mistaken. But from this one can hardly conclude that we have no clear understanding of the concept at all. As we have seen, we have many concepts that we don't know how to analyze. So the mere fact — if indeed it is a fact — that a certain style of analysis of the notion of the physical fails does not mean that there is no notion of the physical at all, still less that we don't understand the notion.

One might object that, while these remarks are perfectly true, they nevertheless don't speak to something that is right about Hempel's dilemma, namely, that for the theory-conception to be complete one needs to know what type of theory a physical theory is. Perhaps one might appeal here to the fact that we have a number of paradigms of what a physical theory is: common sense physical theory, medieval impetus physics, Cartesian contact mechanics, Newtonian physics, and modern quantum physics. While it seems unlikely that there is any *one* factor that unifies this class of theories, perhaps there is a cluster of factors — a common or overlapping set of theoretical constructs, for example, or a shared methodology. If so, one might maintain that the notion of a physical theory is a Wittgensteinian family resemblance concept. However, whether this is enough to answer the question of what kind of theory a physical theory is remains to be seen. (For further discussion of Hempel's dilemma, see the papers in Elpidorou 2018a.)

### 4.4 The panpsychism problem

Hempel's dilemma against the theory-conception is similar to an objection that one often hears propounded against the object-conception (cf. Jackson 1998). Consider a version of panpsychism according to which all the physical objects of our acquaintance are conscious beings just as we are. (For further discussion and different versions of this view, see [Panpsychism](#).) Would physicalism be true in that situation? It seems intuitively not; however, if physicalism is defined via reference to the object-conception of a physical property then it is hard to see why not. After all, according to that conception, something is a physical property just in case it is required by a complete account of paradigmatic physical objects. But this makes no reference to the *nature* of paradigmatic physical objects, and so allows the possibility that physicalism is true in the imagined situation.

One thing to say in response to this objection is that the mere possibility of panpsychism cannot really be what is at issue here. For panpsychism *per se* is not inconsistent with physicalism (cf. Lewis 1983). After all, the fact that there are *some* conscious beings is not contrary to physicalism — why then should the possibility that *everything* is a conscious being be contrary to physicalism? If so, what is at issue in the objection is not panpsychism so much as the possibility that the paradigms or exemplars in terms of which one characterizes the notion of the physical might turn out to be radically different from what we normally assume in a quite specific sense — they might turn out to be in some essential or ultimate respect mental.

Once the problem is put like that, however, the panpsychism problem looks similar to a problem that arises in general whenever one tries to understand or define a concept in terms of paradigmatic objects which fall under it, viz., that these definitions have certain sort of empirical presuppositions that might turn out to be false. Suppose one tried to define the concept red in terms of similarity to paradigmatic red things, such as blood. Pursuing this strategy commits one to the idea that the belief that blood is red is a piece of common knowledge shared among all those who are competent with the term. But that seems wrong — someone who thought that blood was green would be mistaken about blood but not about red. Now this problem is a difficult problem, however — and this is the crucial point for our purposes — the problem is also a quite general problem; it arises because of the paradigm style of definition. So to that extent, the concept of the physical does not seem to be any worse off than the concept of red, the panpsychism problem notwithstanding. (For discussion of the general strategy see Lewis 1997)

Of course, one would reject this entire line of thought if one rejected its starting point, viz., that panpsychism is consistent with physicalism. Wilson (2006, 78–9), for example, suggests that while physicalism is consistent with the view that some conscious beings exist, it is not consistent with the view that some fundamental conscious beings exist, and it is this last claim that is definitive of panpsychism. But in fact even that is consistent with physicalism, though admittedly of an unusual sort. To illustrate, imagine a world in which the fundamental properties are both mental and physical. That is certainly a far-fetched scenario but it doesn't seem to be impossible. Would physicalism be true in such a world? It is hard to see why not; at least it may be true at that world that any physical duplicate of it is a duplicate *simpliciter*. Would panpsychism likewise be true at such a world? Again, it is hard to see why not, since the fundamental properties instantiated at such a world are mental, though of course they are also physical.

## 4.5 The via negativa

One idea that often emerges in the context of Hempel's dilemma and the panpsychism problem, but deserves separate treatment, is the so-called Via Negativa (see e.g. Montero and Papineau 2005, Wilson 2006, Fiorese 2016).

The simplest way to introduce the Via Negativa is to interpret it as a definition of the notion a physical property something like this:  $F$  is a physical property if and only if  $F$  is a non-mental property. But there are many reasons to resist such a definition. Take vitalism. Vitalism isn't true, but it might have been true; there is no contradiction in it for example. So imagine a world in which plants and animals instantiate the key property associated with vitalism, viz., *élan vital*. It seems reasonable to say that in that case plants and animals instantiate a property that is non-physical, i.e. *élan vital* is not physical. And yet one should not say on this account that plants and animals instantiate a mental property, i.e., *élan vital* is not mental. In short, *élan vital* is neither mental nor physical. But the Via Negativa as stated cannot accommodate that fact.

One might try to meet this objection by revising the Via Negativa so that what is intended is only a partial definition along these lines:  $F$  is a physical property only if  $F$  is non-mental. Even so problems remain. As we have seen *élan vital* causes a problem because it is neither mental nor physical. But there might be properties that are both mental and physical. Consider a version of the identity theory according to which being in pain just is c-fibers firing. If we suppose that such a theory is true, is the property of being in pain then mental or physical? Both presumably; but this could not be true on the Via Negativa construed as a definition of what a physical property is, even a partial definition. For if a property is mental and physical, then, given the Via Negativa, it will be both mental and non-mental which (of course) it can't be! Now obviously, there are good questions about whether an identity theory along these lines is or could be true, but regardless of whether it is true, it should not be ruled out simply because of a proposal about how to define the words in which it is stated.

Alternatively, one might try to meet the objection by adopting what Wilson 2006 calls the 'no fundamental mentality' constraint. On this interpretation, what proponents of the Via Negativa have in mind is that  $F$  is a physical property only if  $F$  is not fundamentally mental, where in turn to be 'not fundamentally mental' is most naturally understood as entailing that if  $F$  is a fundamental property then it is non-mental. This version of the view avoids the problem about having c-fibers since presumably that property is not fundamental. But once again problems remain. Take the world we considered above at which the fundamental properties are both mental and physical; in effect, what applies to c-fibers firing (if the identity theory is true) applies to the fundamental properties instantiated at this world. As I said, this scenario is far-fetched, but it doesn't seem to be impossible, and it is certainly not impossible simply as a matter of the definition of the words. And yet it would be impossible for that reason if the 'no fundamental mentality' version of the Via Negativa were true.

Of course, to raise these problems for the Via Negativa is not to deny that there is something right about it. For example, when we think of properties that would falsify physicalism we do often think of \*certain\* mental properties, e.g., the distinctive properties of ectoplasm or ESP. However, this fact—that certain mental properties would, if instantiated, falsify physicalism—can be captured without defining the physical in general as the non-mental. A better way would be to require, of any spelling out of the notion of the physical, whether it be the object-based account or the theory-based account, that it respect that fact that some (uninstantiated) mental properties are non-physical.

## 4.6 Structuralist Approaches

Another idea about how to define the physical that has become prominent in recent times is a structuralist approach to the physical.

One way to introduce the structuralist approach is to see it as a development of the theory conception formulated above. On the theory view, a property is physical just in case it is tied in the right way to a physical theory. One problem that arises for such a view, as we have seen, is what 'physical theory' is supposed to mean here. For the structuralist (at least in the context; 'structuralism' can mean many different things), a physical theory is one that employs a restricted vocabulary. One particularly frank version of this sort of view is suggested by Russell 1927; on this view, the vocabulary in question is restricted to logical or mathematical vocabulary. Contemporary philosophers adopt a less restricted view according to which the vocabulary is either logical or mathematical or causal or nomological (i.e. pertaining to laws) or some combination of these (see Alter 2016, Chalmers 2020, Goff 2017; for criticism see Stoljar 2020). In effect, proposals like this provide a topic-neutral conception of the physical, and hence a topic-neutral conception of physicalism.

An attractive feature of this approach is that it provides an answer to Hempel's dilemma and similar problems. Physics may indeed change over time, but according to the structuralist, any physical theory must be restricted to this sort of vocabulary. Hence we can appeal to the notion of a physical theory to formulate a version of physicalism on which everything supervenes on or is realised by or is grounded in physical properties that can be expressed in that limited vocabulary. (Here structuralism about the physical draws support from structural realism in philosophy of science, a

position to which it is in some ways quite similar; see, e.g. [Structural Realism](#))

However, a problem for structuralism (developed in Stoljar 2020) is that, while placing restrictions of this sort on the notion of a physical theory helps with Hempel's dilemma, it also seems overly stringent. Physical theories on the face of it tell you lots of things about the physical world which are not topic-neutral, for example, about mass, energy, electrons, protons and a myriad other things. The idea that we can capture all of this using a language that employs *only* logical/mathematical or causal/nomic vocabulary may seem an overly ambitious one.

While there are problems with structuralist approaches to the physical, as with the Via Negativa, there is also something right about it. We have noted one idea in philosophy of science that is similar to the structuralist approach to the physical, namely, structural realism. A different idea that philosophers of science have emphasised a lot in recent literature is that physicists and other scientists often construct mathematical models of the systems they are interested in; moreover, they often focus on the mathematical properties of these models themselves (see, e.g. Weisberg 2013). Perhaps a structuralist approach to such mathematical models is plausible. Nevertheless, structuralism about the physical is a thesis, not about these mathematical models, but about the target systems that these models correspond to, and it precisely this that causes the problem.

#### 4.7 Physicalism as an Attitude

In view of the difficulties posed by Hempel's dilemma and related problems, some philosophers have explored the interesting idea that to be a physicalist is not to hold some thesis or belief – that is, to hold something that may be true or false – but is rather to adopt a kind of attitude or stance. As Alyssa Ney (2008, p. 9, see also Van Fraassen 2002) develops this “attitudinal” view, for example, “physicalism is an attitude one takes to form one's ontology completely and solely according to what physics says exists”.

Now, as with other ideas we have looked at, there is certainly something right about the attitudinal view. As we will see below, contemporary physicalists are often methodological naturalists, and methodological naturalists may well hold the attitude Ney describes. Nevertheless, there is a major problem for the view, viz., that on the face of it holding this sort of attitude is neither necessary nor sufficient for being a physicalist.

To see it is not necessary, consider such ancient philosophers as Democritus or Lucretius. These philosophers are physicalists, or at least are usually classified that way, i.e., since they held the doctrine traditionally called ‘materialism’. But they did not hold the attitude Ney describes, either implicitly or explicitly, for physics (at least identified sociologically) did not exist in their day at all.

In response, one might adjust the attitudinal view so that the ‘physics’ towards which one holds the relevant attitude is not identified sociologically, but is instead understood as a certain sort of theory considered in the abstract. But then further problems arise. First, it is now difficult to see the difference between holding the relevant attitude and simply believing a thesis. If one resolves to be guided in one's ontology by the truth of a particular theory, how is that different from just believing the theory? Second, if one holds an attitude toward a particular theory, Hempel's dilemma seems to arise again though in a slightly different form. For which physical theory is meant? If one means current physics, as in fact Ney suggests, then one might argue that this is not an attitude that physicalists should reasonably hold, since current physics is incomplete; and if one means ideal physics, it is hard to see what the content or nature of the attitude is.

To see that the attitude is not sufficient, imagine a situation in which physics postulates properties or objects which are like those postulated by traditional dualists; as Ney puts it imagine “it is the year 3000 AD and physicists have been forced to introduce irreducible mental entities into their theory.” (2008, p. 12). In such a situation, a person might hold the attitude Ney describes, and yet intuitively not be a physicalist.

In response, Ney agrees that this is a possibility but points out, first, it would still be reasonable to criticize the people who hold the attitude – for example, on the grounds that those who hold a different attitude might have arrived at correct ontology more quickly – and, second, that it doesn't follow that the attitude definitive of physicalism is identical to the attitude definitive of dualism. (The ideas underlying this second point are (a) if one adopts the attitudinal view about physicalism then one should in fairness adopt it about dualism as well; and (b) that from the fact that two attitudes coincide in a possible situation it does not follow that they are identical.) However, while both these suggestions might be true, it is hard to see them as responding to the basic point that person who holds the attitude Ney describes in the imagined situation is not correctly described as a physicalist. In principle, after all, such a person may be criticized in many ways; moreover, the fact that holding a particular attitude is not sufficient for being a physicalist does not entail that doing so is necessary for being a dualist.

## 5. The Case Against Physicalism

Having provided an answer to the interpretation question, I now turn to the truth question: is physicalism (as we have interpreted it so far) true? I will first discuss three reasons for supposing that physicalism is not true. Then I will consider the case for physicalism.

### 5.1 Qualia and Consciousness

The main argument against physicalism is usually thought to concern the notion of qualia, the felt qualities of experience. The notion of qualia raises puzzles of its own, puzzles having to do with its connection to other notions such as consciousness, introspection, epistemic access, acquaintance, the first-person perspective and so on. However the idea that we will discuss here is the apparent contradiction between the existence of qualia and physicalism.

Perhaps the clearest version of this argument is Jackson's knowledge argument; see [qualia: the knowledge argument](#). This argument asks us to imagine Mary, a famous neuroscientist confined to a black and white room. Mary is forced to learn about the world via black and white television and computers. However, despite these hardships Mary learns (and therefore knows) all that physical theory can teach her. Now, if physicalism were true, it is plausible to suppose that Mary knows everything about the world. And yet — and here is Jackson's point — it seems she does not know everything. For, upon being released into the world of color, it will become obvious that, inside her room, she did not know what it is like for both herself and others to see colors — that is, she did not know about the qualia instantiated by particular experiences of seeing colors. Following Jackson (1986), we may summarize the argument as follows:

P1. Mary (before her release) knows everything physical there is to know about other people.

P2. Mary (before her release) does not know everything there is to know about other people (because she learns something about them on being released).

Conclusion. There are truths about other people (and herself) that escape the physicalist story.

Clearly this conclusion entails that physicalism is false: for if there are truths which escape the physicalist story how can everything supervene on the physical? So a physicalist must either reject a premise or show that the premises don't entail the conclusion.

There are many possible responses to this argument, but here I will briefly mention only three. The first is *the ability hypothesis* due to Lawrence Nemerow (1988) and developed and defended by David Lewis (1994). The ability hypothesis follows Ryle (1949) in drawing a sharp distinction between propositional knowledge or knowledge-that (such as 'Mary knows that snow is white') and knowledge-how (such as 'Mary knows how to ride a bike'), and then suggests that all Mary gains is the latter. On the other hand, P2 would only be true if Mary gained propositional knowledge.

A second response appeals to the distinction between *a priori* and *a posteriori* physicalism. As we saw above, the crucial claim of a *a posteriori* physicalism is that (13) — i.e. the claim that *S* entails *S\** — is *a posteriori*. Since (13) is *a posteriori*, you would need certain experience to know it. But, it is argued, Mary has not had (and cannot have) the relevant experience. Hence she does not know (13). On the other hand, the mere fact that Mary has not had (and cannot have) the experience to know (13) does not remove the possibility that (13) is true. Hence a *a posteriori* physicalism can avoid the knowledge argument. (It is an interesting question which premise of the knowledge argument is being attacked by this response. The answer depends on whether (13) is physical or not: if (13) is physical, then the response attacks P1. But if (13) is not physical, the response is that the argument is invalid.)

A third response is to distinguish between various conceptions of the physical. We saw above that potentially the class of properties defined by the theory-conception of the physical was distinct from the class of properties defined by the object-conception. But that suggests that the first premise of the argument is open to interpretation in either of two ways. On the other hand, Jackson's thought experiment only seems to support the premise if it is interpreted in the one way, since Mary learns by learning all that physical *theory* can teach her. But leaves open the possibility that one might appeal to the object-conception of the physical to define a version of physicalism which evades the knowledge argument.

One of the most lively areas of philosophy of mind concerns the issue of which if any of these responses to the knowledge argument will be successful. (See the papers in Ludlow, Nagasawa, and Stoljar 2004. See also [Qualia: The Knowledge Argument](#)) The ability response raises questions about whether know-how is genuinely non-propositional (cf. Lycan 1996, Loar 1997 and Stanley and Williamson 2001), and about whether it gets the facts right to begin with (Braddon Mitchell and Jackson 1996). As against a *a posteriori* physicalism, it has been argued both that it rests on a mistaken approach to the necessary *a posteriori* (Chalmers 1996, 1999, Jackson 1998), and that the promise of the idea is chimerical anyway (cf. Stoljar 2000). The third response raises questions about the distinction between the object and the theory conception of the physical and associated issues about dispositional and categorical properties, and also about the relation between physicalism on the one hand, and a related view, sometimes called Russellian monism. See [Russellian monism](#), as well as Chalmers 1996, Lockwood 1992, Stoljar 2000, 2001, Montero 2010, 2015)

## 5.2 Meaning and Intentionality

Philosophers of mind often divide the problems of physicalism into two: first, there are the problems of qualia, typified by the knowledge argument; second, there are problems of intentionality. The intentionality of mental states is their aboutness, their capacity to represent the world as being a certain way. One does not simply think, one thinks *of* (or *about*) Vienna; similarly, one does not simply believe, one believes *that* snow is white. Just as in the case of qualia, some of the puzzles of intentionality derive from facts internal to the notion, and from the relation of this notion to the others such as rationality, inference and language; see [Intentionality](#). But others derive from the fact that it seems difficult to square the fact that mental states have intentionality with physicalism. There are a number of ways of developing this criticism but much recent work has concentrated on a certain line of argument that Saul Kripke has found in the work of Wittgenstein (1982; see also [Private Language](#)).

Kripke's argument is best approached by first considering what is often called a dispositional theory of linguistic meaning. According to the dispositional theory, a word means what it does — for example, the word 'red' means red — because speakers of the word are disposed to apply to word to red things. Now, for a number of reasons, this sort of theory has been very popular among physicalists. First, the concept of a disposition at issue here is clearly a concept that is compatible with physicalism. After all, the mere fact that vases are fragile and sugar cubes are soluble (both are classic examples of dispositional properties) does not cause a problem for physicalism, so why should the idea that human beings have similar dispositional properties? Second, it seems possible to develop the dispositional theory of linguistic meaning so that it might apply also to intentionality. According to a dispositional theory of intentionality, a mental concept would mean what it does because thinkers are disposed to employ the concept in thought in a certain way. So a dispositional theory seems to hold out the best promise of a theory of intentionality that is compatible with physicalism.

Kripke's argument is designed to destroy that promise. (In fact, Kripke's argument is designed to destroy considerably more than this: the conclusion of his argument is a paradoxical one to the effect that there can be no such a thing as a word's having a meaning. However, we will concentrate on the aspects of the argument that bear on physicalism.) In essence his argument is this. Imagine a situation in which (a) the dispositional theory is true; (b) the word 'red' means red for a speaker *S*; and yet (c) the speaker misapplies the word — for example, *S* is looking at a white thing through rose-tinted spectacles and calls it red. Now, in that situation, it would seem that *S* is disposed to apply 'red' to things which are (not merely red but) either-red-or-white-but-seen-through-rose-tinted-spectacles. But then, by the theory, the word 'red' means (not red but) either-red-or-white-as-seen-through-rose-tinted-spectacles. But that contradicts our initial claim (b), that 'red' means red. In other words, the dispositional theory, when combined with a true claim about the meaning of word, plus a truism about meaning — that people can misapply meaningful words — leads to a contradiction and is therefore false.

How might a physicalist respond to Kripke's argument? As with the knowledge argument, there are many responses but here I will mention only two. The first response is to insist that Kripke's argument neglects the distinction between *a priori* and *a posteriori* physicalism. Kripke often does

say that according to the dispositionalist, one should be able to ‘read off’ truths about meaning from truths a physicalist can accept. (For a proposal like this, see Horwich 2000.) One problem with this proposal is, as we have seen, that its background account of the necessary *a posteriori* is controversial. As we saw, *a posteriori* physicalists are committed to what we called the non-derivation view about necessary *a posteriori* truths. But the non-derivation view has come under attack in recent times.

The second response is to defend the dispositional theory against Kripke’s argument. One way to do this is to argue that the argument only works against a very simple dispositionalism, and that a more complicated version of such a theory would avoid these problems. (For a proposal along these lines, see Fodor 1992 and the discussion in Braddon-Mitchell and Jackson 1996). A different proposal is to argue that Kripke’s argument underestimates the complexity in the notion of a disposition. The mere fact that in certain circumstances someone would apply ‘red’ to white things does not mean that they are disposed to apply red to white things — after all, the mere fact that in certain circumstances something would burn does not mean that it is flammable in the ordinary sense. (For a proposal along these lines see Hohwy 1998, and Heil and Martin 1998)

As with the knowledge argument, the issues surrounding Kripke’s argument are very much wide open. But it is important to note that most philosophers don’t consider the issues of intentionality as seriously as the issue of qualia when it comes to physicalism. In different vocabularies, for example, both Block (1995) and Chalmers (1996) distinguish between the intentional aspects of the mind or consciousness, and the phenomenal aspects or qualia, and suggest that it is really the latter that is the central issue. As Chalmers notes (1996; p. 24), echoing Chomsky’s famous distinction, the intentionality issue is a *problem*, but the qualia issue is a *mystery*.

### 5.3 Numbers and Abstracta

A third problem, which we mentioned briefly above, is *the problem of abstracta* (Rabin 2020). This concerns the status within physicalism of abstract objects, i.e., entities apparently not located in space and time, such as numbers, properties and relations, or propositions.

To see the problem, suppose that abstract objects, if they exist, exist necessarily, i.e., in all possible worlds. If physicalism is true, then the facts about such objects must either be physical facts, or else bear a particular relation (grounding, realisation) to the physical. But on the face of it, that is not so. Can one really say that  $5+7=12$ , for example, is realised in, or holds in virtue of, some arrangement of atoms and void? Or can one say that it itself is a physical fact or a fundamental physical fact? If not, physicalism is false: the property of being such that  $5+7=12$  obtains the actual world but is neither identical to, nor grounded in or realized by, any physical property. (Sometimes the problem of abstracta is formulated as concerning, not abstract objects such as numbers or properties, but the grounding or realization facts themselves; see, e.g. Dasgupta 2015. We will set this aside here.)

There are a number of responses to this problem in the literature; for an overview, see Rabin 2020, see also Dasgupta 2015 and Bennett 2017; for more general discussion of physicalism and abstracta, see Montero 2017, Schneider 2017, and Witmer 2017.

One response points out that, while the problem of abstracta confronts many different versions of physicalism, it does not arise for supervenience physicalism. After all, since numbers exist in all possible worlds, facts about them trivially supervene on the physical; any world identical to the actual world in physical respects will be identical to it in respect of whether  $5+7=12$ , because any world at all is identical to the actual world in that respect! But the difficulty here is that supervenience physicalism seems, as we saw above, too weak anyway. Indeed, one might think that the example of abstracta is simply a different way to bring out that it is too weak.

Another option is to adopt a version of nominalism, and deny the existence of abstracta entirely. The problem with this option is that defending nominalism about mathematics is no easy matter, and in any case nominalism and physicalism are normally thought of as distinct commitments.

A third view, which seems more attractive than either of the two mentioned so far, is to expand the notion of a physical property that is in play in formulations of physicalism. For example, one might treat the properties of abstract objects as *topic-neutral* in something like the sense discussed in connection with Smart and reductionism above (see section 3.1). Topic-neutral properties have the interesting feature that, while they themselves are not physical, but are capable of being instantiated in what is intuitively a completely physical world, or indeed what is intuitively a completely spiritual world or a world entirely made of water. If so, it becomes possible to understand physicalism so that the reference to ‘physical properties’ within it is understood more correctly as ‘physical or topic-neutral properties’.

A final response to the problem does not expand the notion of the physical as much as it restricts the scope of physicalism to properties of a certain sort. One suggestion along these lines has been made by a number of writers (e.g. Rabin 2020, following Dasgupta 2015) in the context of grounding physicalism, though perhaps the underlying idea can be extended to other varieties as well. They suggest that grounding physicalism, which we formulated above as (7), should be revised to take the following form:

(7\*) Physicalism is true at a possible world  $w$  iff every *groundable* property instantiated at  $w$  is either a physical property or is grounded in a physical property.

Here, a property is groundable just in case it is apt for being grounded, i.e. it is the sort of property that can be either grounded or not. Ordinary psychological properties are presumably in this class. If grounding physicalism is true, they are grounded in physical properties, whereas if it is false, they are not grounded in physical properties; either way, they are groundable properties. But mathematical properties and the properties of abstracta more generally might in at least some instances *not* be groundable, they may fail to be properties that are apt for being grounded. If so, we have an alternative solution to the problem of abstracta: (7\*) permits that some properties of abstracta are not grounded in the physical, so long as they are ungroundable.

### 5.4 Methodological Issues

The final argument I will consider against physicalism is of a more methodological nature. It is sometimes suggested, not that physicalism is false, but that the entire ‘project of physicalism’ — the project in philosophy of mind of debating whether physicalism is true, and trying to establish or disprove its truth by philosophical argument — is misguided. This sort of argument has been mounted by a number of writers, but perhaps its most

vocal advocate has been Noam Chomsky (2000; see also Searle 1992, 1999).

It is easiest to state Chomsky's criticism by beginning with two points about methodological naturalism. In general it seems rational to agree with the methodological naturalists that the best hope for a theoretical understanding of the world is by pursuing the methods which are typical of the sciences. It would then seem rational as a special case that our best hope for a theoretical understanding of consciousness or experience is by pursuing the methods of the sciences — by pursuing, as we might put it, the naturalistic project with respect to consciousness. So Chomsky's first point is that it is rational to pursue the naturalistic project with respect to consciousness.

Chomsky's second point is that the physicalist project in philosophy of mind is on the face of it rather different from the naturalistic project. In the first place, the physicalist project is, as we have noted, usually thought of a piece of metaphysics. But there is nothing obviously metaphysical about the naturalistic project, it simply raises questions about what we can hope to explain. In the second place, the physicalist project is normally thought of as being amenable to philosophical argument, whereas it is unclear where philosophical argument (if this is different from scientific argument) would enter the naturalistic project. In short, there doesn't seem anything particularly 'philosophical' about the naturalistic project — it simply applies the methods of science to consciousness. But the physicalist project is central to analytic philosophy.

It is precisely at the place where the physicalist project departs from the naturalistic project that Chomsky's criticism begins to take shape. For insofar as it is different from the naturalistic project, there are a number of ways in which the physicalist project is questionable. First, it is hard to see what the project might be — it is true that throughout the history of philosophy and science one encounters suggestions that one might find out about the world in ways that are distinct from the ones used in the sciences, but these suggestions have always been rather obscure. Second, it is hard to see how this sort of project could recommend itself to physicalists *themselves* — such a project seems to be a departure from methodological naturalism but most physicalists endorse methodological naturalism as a matter of fact. On the other hand, if the physicalist project does not depart from the naturalistic project, then the usual ways of talking and thinking about that project are highly misleading. For example, it is misleading to speak of it as a piece of metaphysics as opposed to a piece of ordinary science.

In sum, Chomsky's criticism is best understood as a kind of dilemma. The physicalist project is either identical to the naturalistic project or it is not. If it is identical, then the language and concepts that shape the project are potentially extremely misleading; but if it is not identical, then there are a number of ways in which it is illegitimate.

How is one to respond to this criticism? In my view, the strongest answer to Chomsky accepts the first horn of his dilemma and suggests that what philosophers of mind are really concerned with is the naturalistic project. Now, of course, what concerns them is not so much the details of the project — that would not distinguish them from working scientists. Rather they are concerned with what the nature of the project and what its potential limits might be.

This theme might be developed in several ways, but one well-known development of it has been suggested by Thomas Nagel (1983) and Bernard Williams (1985). According to them, any form of scientific inquiry will at least be objective, or will result in an objective picture of the world. On the other hand, we have a number of arguments — the most prominent being the knowledge argument — which plausibly show that there is no place for experience or qualia in a world that is described in purely objective terms. If Nagel and Williams are right that any form of scientific inquiry will yield a description of the world in objective terms, the knowledge argument is nothing less than a negative argument to the effect that the naturalistic project with respect to consciousness will not succeed.

If what is at issue is the limits of the naturalist project, why is the debate so often construed as a metaphysical debate rather than a debate about the limits of inquiry? In answer to this question, we need to sharply divorce the background metaphysical framework within which the problems of philosophy of mind find their expression, and the problems themselves. Physicalism is the background metaphysical assumption against which the problems of philosophy of mind are posed and discussed. Given that assumption, the question of the limits of the naturalistic project *just is* the question of whether there can be experience in a world that is totally physical. Nevertheless, when properly understood, the problems that philosophers of mind are interested in are not with the framework themselves, and to that extent are not metaphysical. Thus, the common phrase 'metaphysics of mind' is misleading.

## 6. The Case for Physicalism

Having considered one side of the truth question, I will turn finally to the other: what reason is there for believing that physicalism is true?

The first thing to say when considering the truth of physicalism is that we live in an overwhelmingly physicalist or materialist intellectual culture. The result is that, as things currently stand, the standards of argumentation required to persuade someone of the truth of physicalism are much lower than the standards required to persuade someone of its negation. (The point here is a perfectly general one: if you already believe or want something to be true, you are likely to accept fairly low standards of argumentation for its truth.)

However, while it might be difficult to assess dispassionately the arguments for or against physicalism, this is still something we should endeavor to do. Here I will review two arguments that are commonly thought to establish the truth of physicalism. What unites the arguments is that each takes something from the physicalist world-picture which we considered previously and tries to establish the metaphysical this of physicalism.

The first argument is (what I will call) *The Argument from Causal Closure*. The first premise of this argument is the thesis of the Causal Closure of the Physical — that is, the thesis that every event which has a cause has a physical cause. The second premise is that mental events cause physical events — for example we normally think that events such as wanting to raise your arm (a mental event) cause events such as the raising of your arm (a physical event). The third premise of the argument is a principle of causation that is often called the exclusion principle (Kim 1993, Yablo 1992, Bennett 2003). The correct formulation of the exclusion principle is a matter of some controversy but a formulation that is both simple and plausible is the following:

### Exclusion Principle

If an event  $e$  causes event  $e^*$ , then there is no event  $e\#$  such that  $e\#$  is non-supervenient on  $e$  and  $e\#$  causes  $e^*$ .



The conclusion of the argument is the mental events are supervenient on physical events, or more briefly that physicalism is true. For of course, if the thesis of Causal Closure is true then behavioral events have physical causes, and if mental events also cause behavioral events, then they must supervene on the physical if the exclusion principle is true.

The Argument from Causal Closure is perhaps the dominant argument for physicalism in the literature today. But it is somewhat unclear whether it is successful. (For some discussion see, [Mental Causation](#)). One response for the anti-physicalist is to reject the second premise and to adopt a version of what is called epiphenomenalism, the view that mental events are caused by, and yet do not cause, physical events. The argument against this position is usually epistemological: if pains don't cause pain behavior how can it be that your telling me that you are in pain gives me any reason for supposing you are? It might seem that epiphenomenalists are in trouble here, but as a number of recent philosophers have argued, the issues here are very far from being settled (Chalmers 1996, Hyslop 1999). The crucial point is that the causal theory of evidence is open to serious counterexamples so it is unclear that it can be used against epiphenomenalism effectively.

A different sort of response is to reject the causal principles on which the argument is based. As against the exclusion principle, for example, it is often pointed out that certain events are overdetermined. The classic example is the firing squad: both the firing by soldier A and by soldier B caused the prisoner's death but since these are distinct firings, the exclusion principle is false. However, while this line of response is suggestive, it is in fact rather limited. It is true that the case of the firing squad represents an exception to the exclusion principle — an exception that the principle must be emended to accommodate. But is difficult to believe that it represents an exception that can be widespread. A more searching response is to reject the very idea of causal closure on the grounds, perhaps, that (as Bertrand Russell (1917) famously argued) causation plays no role in a mature portrayal of the world. Once again, however, the promise of this response is more imagined than real. While it is true that many sciences do not explicitly use the notion of causation, it is extremely unlikely that they do not imply that various causal claims are true.

The second argument for physicalism is (what I will call) *The Argument from Methodological Naturalism*. The first premise of this argument is that it is rational to be guided in one's metaphysical commitments by the methods of natural science. Lying behind this premise are the arguments of Quine and others that metaphysics should not be approached in a way that is distinct from the sciences but should rather be thought of as continuous with it. The second premise of the argument is that, as a matter of fact, the metaphysical picture of the world that one is led to by the methods of natural science is physicalism. The conclusion is that it is rational to believe physicalism, or, more briefly that physicalism is true.

The Argument from Methodological Naturalism has received somewhat less attention in the literature than the Argument from Causal Closure. But it seems just as persuasive — in fact, rather more so. For how might one respond? One possibility is to reject its first premise. But this is not something that most people are attracted to (or at least are attracted to explicitly.)

The other possibility is to reject its second premise. However, if physicalism can be clearly stated — admittedly, this is a big 'if' — it is not terribly clear what this would amount to or what the motivation for it would be. In the first place, our earlier discussion shows that physicalism is not inconsistent with explanatory autonomy of the various sciences, so that one should not reject physicalism merely because one can't see how to reduce those sciences to others. In the second place, while it is perfectly true that there are examples of non-physicalist approaches to the world — vitalism in biology is perhaps the best example — this is beside the point. The second premise of the Argument from Methodological Naturalism does not deny that other views are possible, it simply says that physicalism is the most likely view at the moment. Finally, one might be inclined to appeal to arguments such as the knowledge argument to show that physicalism is false, and hence that methodological naturalism could not show that physicalism is false. However, this suggestion represents a sort of confusion about the knowledge argument. As we saw above, if successful the knowledge argument suggests, not simply that physicalism is false but that any approach to the world that is compatible with methodological naturalism is false. But if that is so, it is mistaken to suppose that the knowledge argument gives one any reason to endorse anti-physicalism if that is supposed to be a position compatible with methodological naturalism.

## Bibliography

- Alter, T., 2016, 'The Structure and Dynamics Argument against Materialism', *Noûs*, 50(4): 794–815.
- Armstrong, D., 1968, *A Materialist Theory of the Mind*, London: Routledge.
- Bacon, J., 1990, 'Van Cleve Versus Closure', *Philosophical Studies*, 58: 239–242.
- Baltimore, J., 2013, 'Stoljar's Twin-physics World', *Philosophia*, 41: 127–136.
- Bennett, K., 2003, 'Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it', *Noûs*, 37(3): 471–497.
- , 2017, *Making Things Up*, Oxford: Oxford University Press.
- Berkeley, G., 1710, *Principles of Human Knowledge*, London: Penguin, 2004.
- Berker, S., 2018, 'The Unity of Grounding', *Mind*, 127(507): 729–777.
- Blackburn, S., 1992, 'Filling in Space', *Analysis*, 50(2): 60–65.
- Block, N., 1980, 'Troubles with Functionalism', in N. Block (ed.), *Readings in the Philosophy of Psychology*, Volume 1, Cambridge, MA: Harvard University Press, 1980, pp. 268–306.
- , 1995, 'On a Confusion About a Function of Consciousness', *Behavioral and Brain Sciences*, 18(2): 227–287.
- Block, N., and Stalnaker, R., 1999, 'Conceptual Analysis, Dualism and the Explanatory Gap', *Philosophical Review*, 108(1): 1–46.
- Boyd, R., 1980, 'Materialism Without Reductionism: What Physicalism Does not Entail', in N. Block (ed.), *Readings in the Philosophy of Psychology* Volume 1, Cambridge, MA: Harvard University Press, 1980, pp. 76–106.
- Braddon-Mitchell, D., and Jackson, F., 1996, *Philosophy of Mind and Cognition*, Oxford: Blackwell.
- Broad, C.D., 1925, *The Mind and Its Place in Nature*, London: Routledge and Kegan Paul.
- Byrne, A., 1993, *The Emergent Mind*, Ph.D. Dissertation, Philosophy Department, Princeton University.
- , 1999, 'Cosmic Hermeneutics', in J. Tomberlin (ed.), *Philosophical Perspectives*, 13: 347–83.
- Carnap, R., 1932/33 'Psychology in Physical Language', in A.J. Ayer (ed.), *Logical Positivism*, New York: The Free Press, 1959, pp. 165–198.
- Chalmers, D., 1996, *The Conscious Mind*, New York: Oxford University Press
- , 1999, 'Materialism and the Metaphysics of Modality', *Philosophy and Phenomenological Research*, 59: 473–493.
- , 2002, *Philosophy of Mind: Contemporary and Classical Readings*, New York: Oxford University Press
- , 2002a, 'Consciousness and its Place in Nature', in Chalmers 2002, pp. 102–142.