

PDF version of the entry
Physicalism
<https://plato.stanford.edu/archives/win2017/entries/physicalism/>
from the WINTER 2017 EDITION of the

STANFORD ENCYCLOPEDIA OF PHILOSOPHY



Edward N. Zalta Uri Nodelman Colin Allen R. Lanier Anderson
Principal Editor Senior Editor Associate Editor Faculty Sponsor

Editorial Board
<http://plato.stanford.edu/board.html>

Library of Congress Catalog Data
ISSN: 1095-5054

Notice: This PDF version was distributed by request to members of the Friends of the SEP Society and by courtesy to SEP content contributors. It is solely for their fair use. Unauthorized distribution is prohibited. To learn how to join the Friends of the SEP Society and obtain authorized PDF versions of SEP entries, please visit <https://leibniz.stanford.edu/friends/>.

Stanford Encyclopedia of Philosophy
Copyright © 2017 by the publisher
The Metaphysics Research Lab
Center for the Study of Language and Information
Stanford University, Stanford, CA 94305

Physicalism
Copyright © 2017 by the author
Daniel Stoljar

All rights reserved.

Copyright policy: <https://leibniz.stanford.edu/friends/info/copyright/>

Physicalism

First published Tue Feb 13, 2001; substantive revision Mon Mar 9, 2015

Physicalism is the thesis that everything is physical, or as contemporary philosophers sometimes put it, that everything supervenes on the 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 nature. But they insist nevertheless that at the end of the day such items are either physical or supervene on the physical.

- 1. Terminology
- 2. A Framework for Discussion
- 3. Supervenience Physicalism: Introductory
- 4. Supervenience Physicalism: Further Issues
 - 4.1 The lone ammonium molecule problem
 - 4.2 The modal status problem
 - 4.3 The epiphenomenal ectoplasm problem
 - 4.4 The blockers problem
- 5. Supervenience Physicalism as Minimal Physicalism
- 6. Token and Type Physicalism
- 7. Reductive and Non-Reductive Physicalism
- 8. *A Priori* and *A Posteriori* Physicalism
- 9. Is Supervenience Sufficient for Physicalism?
- 10. Non-modal Definitions of Physicalism
 - 10.1 Second-order properties

- 10.2 Subsets of causal powers
 - 10.3 Grounding
 - 11. Understanding ‘Physical’: Introductory
 - 12. Understanding ‘Physical’: Further Issues
 - 12.1 Circularity
 - 12.2 Hempel's dilemma
 - 12.3 The panpsychism problem
 - 12.4 The via negativa
 - 12.5 Physicalism as an Attitude
 - 13. Physicalism and the Physicalist World Picture
 - 14. The Case Against Physicalism I: Qualia and Consciousness
 - 15. The Case Against Physicalism II: Meaning and Intentionality
 - 16. The Case Against Physicalism III: Methodological Issues
 - 17. The Case for Physicalism
 - 18. Further Questions
 - Bibliography
 - Academic Tools
 - Other Internet Resources
 - Related Entries
-

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’ is very old, 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. It is not clear that Neurath and Carnap understood physicalism in the same way, but 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 (i.e. materialism) is associated with a number of other metaphysical and methodological doctrines. We will return to some of these when we discuss Physicalism and the Physicalist World-picture.

2. 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 we will begin with the interpretation question. Nevertheless, the issues here turn out to be somewhat technical, and those new to the topic might like to read only the first section of our discussion of the interpretation question, which is: Supervenience Physicalism: Introductory, and then turn directly to the truth question which begins at The Case Against Physicalism I: Qualia and Consciousness.

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

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

In other words, the completeness question holds fixed the issue of what it means for something to satisfy the condition of being physical, and asks instead what it means for *everything* to satisfy that condition. Notice 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?

The condition question asks:

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

In other words, the condition question holds fixed the issue of what it means for everything to satisfy some condition or other, and asks instead what *is* the condition, being physical, that everything satisfies. Notice again that a parallel question could be asked of Thales: assuming we know what it is for everything to satisfy some condition or other, what is the condition, being water, that according to Thales, everything satisfies? In discussing the interpretation question, I will turn first to the completeness question, and then consider the condition question.

3. Supervenience Physicalism: Introductory

One answer to the completeness question – an answer suggested initially in Davidson 1970 – looks to the notion of supervenience. (This notion is historically associated with meta-ethics, but it has received extensive discussion in the general metaphysics and logic literature. For a survey, see supervenience.) As we shall see as we proceed, supervenience is not

quite as popular an answer to the completeness question as it once was, but it is nevertheless remains an attractive answer, and we may go a considerable distance by concentrating on it.

The idea of supervenience 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).

Lewis's example gives us one way to introduce 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 are nothing but a pattern in the dots, so too the psychological, the biological and the social features of the world are nothing but a pattern in the physical features of the world. To use the language of supervenience, 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. Lewis 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 put this slightly differently, we might

say that if physicalism is true at our world, then no *other* world can be physically identical to it without being identical to it in all respects. This suggests the following 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), then we have an answer to the completeness question. The completeness question asks: what does it mean to say that *everything* is physical. According to (1), what this means is that if physicalism is true, there is no possible world which is identical to the actual 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*.

4. Supervenience Physicalism: Further Issues

Supervenience physicalism is relatively simple and clear, but when construed as a formulation of physicalism, it faces four problems. (There is also a fifth problem to be discussed later, viz., whether supervenience is sufficient for physicalism.) These are: (a) the lone ammonium molecule problem; (b) the modal status problem; (c) the epiphenomenal ectoplasm problem; and (d) the blockers problem. As we will see, while some of these are relatively easily answered, others are much more challenging.

4.1 The lone ammonium molecule problem

(Cf. Kim 1993.) Imagine a possible world W^* that is physically exactly like our world except in one trivial respect: it has one extra ammonium

molecule located, say, on Saturn's rings. It is natural to suppose that at W^* , the distribution of mental properties is exactly as it is in the actual world — the presence of an extra molecule does not make that much of a difference. On the other hand, for all (1) says, such a world might be *radically different* in terms of the distribution of mental properties. Since (1) only refers to worlds that are exact minimal physical duplicates of our world, it is silent on worlds that are different *even in minute details*, and hence is silent on W^* . It thus leaves open the possibility that, in W^* , everything has mental properties, or nothing has or the only things that have mental properties are Saturn's rings! But that seems absurd: while W^* is clearly not a physical duplicate of our world, for it contains an extra molecule, the distribution of mental properties in W^* would nevertheless match that of the actual world.

There are a number of different responses to this problem in the literature (cf. Kim 1993). Perhaps the simplest response is that the problem conflates two issues that are better kept apart: the question of what physicalism itself tells us about W^* , and the question of what our general knowledge tells us about W^* . It is true that physicalism itself does not tell us anything about the distribution of mental properties at W^* . Nevertheless, we know independently what the distribution is — we know independently that the presence or absence of molecules on Saturn doesn't affect things like who has mental properties here on Earth. But why should one assume that this last piece of knowledge should be a consequence of physicalism? To put the point slightly differently, imagine that we discover that who has mental properties on Earth *is* in part a function of the behavior of molecules on Saturn. That would of course tell us that we are deeply wrong in our assumptions about how the world works. But it would not tell us that we are deeply wrong about physicalism. (For further discussion of this point, see Paull and Sider 1992, and Stalnaker 1996.)

4.2 The modal status problem

Some philosophers (e.g. Davidson 1970) have thought of physicalism as a conceptual or necessary truth, if it is true at all. But most have thought of it as contingent, a truth about our world which might have been otherwise. The statement of physicalism encoded in (1) allows a way in which this might be so. (1) tells us that physicalism is true at a world just in case the world in question conforms to certain conditions. But it leaves it open whether or not the actual world conforms to those conditions *as a matter of fact*. Perhaps it is *not* true of our world that a physical duplicate of it would be a psychological duplicate. If so, physicalism would not be true at our world.

But for some it is puzzling that physicalism is stated using modal notions (i.e. notions such as possible worlds) and nonetheless is contingent. To see the problem, notice first that, supervenience physicalism tells us that the physical truths of the world entail *all* the truths; hence

(2) 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.) If supervenience physicalism is true, it will then be true that:

(3) S entails S^*

On the other hand, (3) is clearly a necessary truth. However, if (3) is a necessary truth, how can *physicalism* be contingent? After all, (3) seems equivalent to physicalism. But if the two are equivalent, how can one be necessary and the other contingent?

But the response to this problem is straightforward. (3) is necessary, but it is *not* equivalent to physicalism. Rather, (3) *follows* from physicalism

given various contingent assumptions, in particular the assumptions that S and S^* are the statements we say they are — it is contingent fact, for example that S^* summarizes the total nature of the world. On other hand, (2) is equivalent to physicalism but it is *not* necessary. (It is important to bear in mind here that not all entailment claims are necessary. Consider ‘my aunt's favorite statement entails my uncle's favorite’ — that statement is contingent even though it is most naturally thought of as an entailment claim.)

4.3 The epiphenomenal ectoplasm problem

(Cf. Horgan 1983, Lewis 1983.) Imagine a possible world W that is exactly like our world in respect of the distribution of physical and mental properties, but for one difference: it contains some pure experience which does not interact causally with anything else in the world — *epiphenomenal ectoplasm*, to give it a name. The problem this possibility presents for (1) is that, if (1) provides the correct definition of physicalism, and if physicalism is true at the actual world, then there is *no* possible world of the kind we just described, i.e., W does not exist. The reason is that W is by assumption a physical duplicate of our world; but then, if physicalism is true at our world, W should be a duplicate simpliciter of our world. But W is patently not a duplicate of our world: it contains some epiphenomenal ectoplasm that our world lacks. On the other hand, it seems quite wrong to say that W is an impossibility — at any rate, physicalism should not *entail* that it is impossible.

In order to solve the epiphenomenal ectoplasm problem, we need to adjust (1) so that it does not have the truth of physicalism ruling out W as a possible world. While there are a number of different proposals about how to do this, one influential proposal is due to Frank Jackson (cf. Jackson 1993. For earlier proposals and further discussion, see Horgan 1983 and Lewis 1983.) He proposes replacing (1) with:

(4) Physicalism is true at a possible world w iff any world which is a minimal physical duplicate of w is a duplicate of w *simpliciter*

By ‘minimal physical duplicate’, Jackson means a possible world that is identical in all physical respects to the actual world, but which does not contain anything else; in particular, it does not contain any epiphenomenal ectoplasm. Unlike (1), (4) does not have physicalism ruling out W , and so (4) is on the face of it preferable to (1) as a statement of physicalism.

A different proposal is due to David Chalmers (1996). He proposes replacing (1) with:

(5) Physicalism is true at a possible world w iff any world which is a physical duplicate of w is a positive duplicate of w

By ‘positive duplicate’, Chalmers means a possible world that instantiates all the positive properties of the actual world, where in turn a positive property is defined as “one that if instantiated in a world W , is also instantiated by the corresponding individual in all worlds that contain W as a proper part” (1996, p. 40). Unlike (1), and like (4), (5) does not have physicalism ruling out W , and so (5) is on the face of it preferable to (1) as a statement of physicalism.

4.4 The blockers problem

(Cf. Hawthorne 2002, Leuenberger 2008) Imagine a possible world similar to ours with respect to the distribution of mental and physical properties, except for this difference: the relation between physical facts and mental facts is weaker than supervenience—mental facts are entailed by physical facts so long as there are no facts which block that entailment—blockers, as they are called. For example, being in an overall physical condition P will necessitate being in pain so long as you do not also

instantiate some further property B. If you are in both P and B you are not in pain; but if you are in P and not in B, you will be in pain.

The problem that this possibility raises for supervenience definitions of physicalism is as follows. Let us suppose that the relation obtaining at a world *W* between the mental and the physical is one of weak necessity as just defined; that is, suppose that, at *W*, the mental is necessitated by the physical but only if certain blockers are absent. Intuitively it would seem that physicalism is false at *W*. On the other hand, if physicalism is defined in the way suggested by Jackson it would be true. After all, applied to *W*, Jackson's definition says that physicalism is true at *W* just in case any minimal physical duplicate of *W* is a duplicate simpliciter. But that seems to be true of *W* as we have imagined it. Conclusion: if blockers are possible, physicalism is false at *W*, and yet it should not be false on Jackson's definition.

There are a number of possible responses to the blockers problem. One is to resist the intuition that physicalism is false at *W* in the circumstance described, even if we adopt Jackson's definition of physicalism. A different response is to adopt a formulation of physicalism that is weaker than supervenience physicalism; this strategy is pursued in Leuenberger 2008. A third response is to say that what the blockers problem brings out is that there is a difference between two ways that physicalists have sought to respond to the epiphenomenal ectoplasm problem; in particular, if one adopts (5) rather than (4) as one's response to the epiphenomenal ectoplasm problem, this would have the advantage that it does not also face the blockers problem. For if the relation of the mental to the physical that obtains at *W* is one of weak necessity, then not only is physicalism false but it is also false that any world which is a physical duplicate of *W* is a positive duplicate of *W* — at some physical duplicate worlds, for example, there will be no psychological properties at all. None of these responses are clearly correct however, and the proper treatment of the

blockers problem (and indeed of the epiphenomenal ectoplasm problem, of which the blockers problem is a development) is an open question in the literature (For simplicity, I will continue with (1), rather than with either (4) or (5); nothing in what follows will turn on this choice.)

5. Supervenience Physicalism as Minimal Physicalism

Physicalism is intended as a very general claim about the nature of the world, but by far the most discussion of physicalism in the literature has been in the philosophy of mind. The reason for this is that it is in philosophy of mind that we find the most plausible and compelling arguments that physicalism is false. Indeed, as we will see later on, arguments about qualia and consciousness are usually formulated as arguments for the conclusion that physicalism is false.

While the issue of physicalism is central to philosophy of mind, however, it is important also to be aware that supervenience physicalism is neutral on a good many of the questions that are pursued in philosophy of mind, and pursued elsewhere for that matter. If you read over the philosophy of mind literature, you will often find people debating a number of different issues: whether there *are* mental states at all; what *sort* of thing mental states are; to what extent mental states are environmentally determined. Given the multifariousness of mental states, it is quite likely that the correct position will be some kind of combination of these positions. But this is a question of further inquiry that is irrelevant to physicalism itself. So physicalism itself leaves many debates in the philosophy of mind unanswered.

This point is sometimes expressed by saying that *supervenience physicalism is minimal physicalism* (Lewis 1983): it is intended to capture the minimal or core commitment of physicalism. Physicalists may differ

from one another in many ways, but all of them must at least hold supervenience physicalism. (Notice that the idea that (1) captures the minimal commitment of physicalism is a distinct idea from that of a minimal physical duplicate which Jackson uses in his attempt to capture minimal physicalism.)

Two issues here require further comment. First, in some discussions in philosophy of mind, the term ‘physicalism’ is used to refer to the identity theory, the idea that mental states or properties are neurological states or properties (Block 1980). In this use of the term, one can reject physicalism by rejecting the identity theory — so by that standard a behaviorist or functionalist in philosophy of mind would not count as a physicalist. Obviously, this is a much more restricted use of the term than is being employed here.

Second, one might think that supervenience physicalism is inconsistent with eliminativism, the claim that psychological states do not exist, for the following reason. Suppose psychological states supervene on physical states. Doesn't that mean, contrary to eliminativism, that there must *be* some psychological states? The answer to this question is ‘no.’ For consider: the telephone on my desk has no psychological states whatsoever. Nevertheless it is still true (though, admittedly, a little odd) to say that a telephone which is identical to my telephone in all physical respects will be identical to it in all psychological respects. In the sense intended, therefore, one thing can be psychologically identical to another even when neither *has* any psychological states.

6. Token and Type Physicalism

To what extent does supervenience physicalism capture minimal physicalism, the core commitment of all physicalists? In order to answer this question it is worth comparing and contrasting supervenience

physicalism with two alternative statements of physicalism that one finds in the literature: token and type physicalism.

Token physicalism is the view that every particular thing in the world is a physical particular. Here is one formulation of this idea:

Token physicalism:

For every actual particular (object, event or process) x , there is some physical particular y such that $x = y$.

Supervenience physicalism neither implies nor is implied by token physicalism. To see that token physicalism does not imply supervenience physicalism, one need only note that the former is consistent with a version of dualism, namely property dualism. The mere fact that every particular has a physical property does not rule out the possibility that some particulars also have non-supervenient mental properties, i.e. mental properties that are only contingently related to the physical. But supervenience physicalism *does* rule out this possibility. Since token physicalism does not rule out property dualism but supervenience physicalism does, the first does not imply the second.

To see that supervenience physicalism does not imply token physicalism, consider the United States Court of Appeals for the Seventh Circuit. This might be thought of as a social or legal object. But then, according to token physicalism, there must be some physical object for it to be identical with. But there might be no physical object (in any natural sense of the term) which is identical to the US Court of Appeals for the Seventh Circuit. On the other hand, supervenience physicalism imposes no such requirement, and so supervenience physicalism does not imply token physicalism (For the classic presentation of this point, see Haugeland 1983).

The point that supervenience physicalism is logically distinct from token physicalism is an important one. One thing it shows is that token physicalism (since it is consistent with property dualism) does not capture minimal physicalism, and so the distinction between token physicalism and supervenience physicalism is no objection to the latter. But the difference between the two theses also raises a different question. Given that token physicalism does not capture the minimal commitment of physicalism, why has token physicalism been the subject of such discussion? One reason is that token physicalism provides one version of the idea that upper level scientific claims requires physical mechanisms. Supervenience physicalism does not on its own entail this. But token physicalism is often seen as a way to ensure this requirement. (For the classic presentation of this point, see Fodor 1974; see also Papineau 1996. And for a different view about token physicalism see Latham 2001.)

Having considered token physicalism, we can now turn to type physicalism. Type physicalism is a generalization and extension of the identity theory, which we considered above. It holds that that every property (or at least every property that is or could be instantiated in the actual world) is identical with some physical property. Here is a statement of this sort of idea:

Type physicalism:

For every actually instantiated mental property F , there is some physical property G such that $F=G$.

Unlike token physicalism, type physicalism certainly *does* entail supervenience physicalism: if every property instantiated in the actual world is identical with some physical property, then a world identical to our world in physical respects will of course be identical to it in all respects.

Nevertheless the reverse entailment does not hold. Supervenience physicalism, as we have been understanding it, is consistent with the possibility (if not the actuality) of disembodiment. But type physicalism as defined here is inconsistent with this possibility, at least if we focus actually instantiated mental properties. To that extent, supervenience physicalism does not entail type physicalism.

Earlier we noted that philosophers such as Davidson have thought that physicalism is a necessary truth. Even on that assumption, however, it is still not completely obvious that supervenience physicalism entails type physicalism. 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.)

7. Reductive and Non-Reductive Physicalism

Before the development of the notion of supervenience, physicalism was often stated as a reductionist thesis. It will therefore be useful to contrast the supervenience formulation of physicalism with various reductionist proposals, and also to consider a question that has received a lot of attention in the literature, viz., whether a physicalist must be a reductionist.

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 usually 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 (6):

(6) 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 (6) in fact, no physicalist since before Smart (1959) has (unqualifiedly) held anything like (6). 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 (6). It is fair to say that this move is one of the central innovations of philosophy of mind, a move to a large extent endorsed and developed later on by functionalists and cognitive scientists.

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:

(7) Reductionism is true iff for each mental predicate F there is a neurological 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 (7). Indeed, many philosophers have argued that there are very strong empirical reasons to deny that anything like (7) is going to be the case. The reason is this. Many different neurological processes (whether in our own species or a different one) could underlie the same psychological process — indeed, given science fiction, even non-neurological processes might underlie the same psychological process. But if multiple realizability — as this sort of idea is called — is true, then (7) seems to be false. (Fodor 1974, but for recent alternative views, 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 neurological theory — in other words, this version of reductionism is in essence a version of type physicalism or the identity theory. However, as we have seen, if physicalists are committed only to supervenience physicalism, they are not committed to type physicalism. Hence a physicalist need not be a reductionist in this metaphysical sense.

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,

(8) 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 (8) says is that if reductionism is true, *a priori* knowledge alone, plus knowledge of the physical truths will allow one to know the mental truths. This question is in fact a highly vexed one in contemporary philosophy. However, this question is usually debated in the context of another, viz., the question of *a posteriori* and *a priori* physicalism. It is to that question, therefore, to which we will now turn.

8. *A Priori* and *A Posteriori* Physicalism

We saw earlier that if physicalism is true then (3) is true, where ‘ S ’ is a sentence that reports the entire physical nature of the world and ‘ S^* ’ is a sentence that reports the entire nature of the world:

(3) S entails S^*

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

(9) If S then S^*

Indeed, this is a general feature of physicalism: if it is true then there will always be a necessary truth of the form of (9).

Now, if (9) 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 (9) 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 in a moment. Indeed, 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 (9) 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 (9). However, if $S\#$ logically entails (9) 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*:

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

One the other hand, if physicalism is true, and S summarizes the total nature of the world it seems reasonable to suppose that $S\#$ was *already implicitly included in S*. In other words it seems reasonable to suppose that (10) is simply an expansion of (9). But if (10) is just an expansion of (9), then if (10) is *a priori*, (9) must also be *a priori*. But that means our initial assumption is false: (9) is not a necessary *a posteriori* truth after all (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. Indeed, the question of which interpretation of Kripke's work is the right one, is one of the most vexed in contemporary analytic philosophy. 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)

9. Is Supervenience Sufficient for Physicalism?

We noted above that while supervenience provides an attractive answer to the completeness question, it is not as popular now as it once was. Part of the reason for this are the problems mentioned in Supervenience Physicalism: Further Issues. But perhaps the most influential consideration here is what I will call *the sufficiency problem*, viz., the apparent fact that supervenience provides a necessary but not sufficient condition for the truth of physicalism.

One way to bring out the sufficiency problem focuses on *emergentism*, a position on the mind-body problem influential in the first forty years of the

twentieth century (Cf. Kim 1998; see also Wilson 2005; for the historical background to emergentism, see MacLaughlin 1992.) Emergentism may itself be understood in several ways, but in the sense that matters to the sufficiency objection, what is intended is a position that weaves together elements of both dualism and physicalism. On the one hand, the emergentist wants to say that mental facts and physical facts are metaphysically distinct—just as a standard dualist does. On the other hand, emergentist wants to agree with the physicalist that mental facts are necessitated by, and so supervene on, the physical facts. If this sort of position is coherent, (1) does not articulate a sufficient condition for physicalism. For if emergentism is true, any physical duplicate of the actual world is a duplicate *simpliciter*. And yet, if emergentism is true, physicalism is false.

A different way to bring out the sufficiency problem focuses on the idea of a necessary being which is essentially nonphysical (Cf. Jackson 1998). (Some theists believe that God provides an example of such a being.) If such a non-physical being exists, it is natural to suppose that physicalism is false. But if physicalism is defined according to (1), then physicalism may still be true, for it remains possible that any minimal physical duplicate of the world is a duplicate *simpliciter*. So, again, (1) does not formulate a sufficient condition for the truth of physicalism.

How to respond to the sufficiency problem? Some philosophers suppose that the issue is so serious that the only thing to do is to retreat from supervenience physicalism to type physicalism (e.g. Kim 1998). The major burden on this proposal, however, is that, as we have seen, type physicalism was given up for a very good reason, e.g., multiple realizability.

A different suggestion points out that the problem is only genuine if the cases that generate it are coherent – and are they? One reason against

supposing so is that both seem to violate Hume's dictum that there are no necessary connections between distinct existences. According to emergentism, for example, mental and physical properties are metaphysically distinct, and yet are necessarily connected. And if the non-physical necessary being exists at all, it will presumably be necessarily connected to the physical world and yet distinct from it. However, Hume's dictum is itself a matter of controversy, so it is unclear if the cases can be dismissed in this way (see Jackson 1993, Stalnaker 1996, Stoljar 2010, and Wilson 2005, 2010).

In view of the difficulty of responding to the sufficiency problem by either retreating to type physicalism or rejecting the examples that generate the problem, a natural option at this point is to agree that (1) does not articulate a sufficient condition for physicalism but search for a related proposal that does. One suggestion along these lines, for example, might be to replace (1) with:

(11) Physicalism is true at a possible world w iff (a) any world which is a physical duplicate of w is a duplicate of w simpliciter; and (b) no property instantiated at w is metaphysically distinct from the physical properties instantiated at w .

Unlike (1), (11) does not have the consequence that the supervenience is sufficient for physicalism; hence it does not entail that physicalism is true if either emergentism is true or essentially non-physical necessary beings exist. Of course, (11) faces the further problem of saying what metaphysical distinctness is, and more generally what the relation is between condition (a) and condition (b). But it might also be supposed on behalf of (11) that the proponent of the sufficiency objection must already have answers to these questions, for otherwise the objection could not itself be advanced.

10. Non-modal Definitions of Physicalism

We have looked at the sufficiency objection and noted one way in which the supervenience approach might meet it. But one might think that the problem goes deeper than has been brought out so far. According to this line of thought, what is lying behind the sufficiency objection is the fact that (1) is a *modal* definition of physicalism, i.e. one in terms of necessity and possibility. But, the suggestion continues, modal definitions inevitably face problems of this sort. If so, the proper response to the objection is to not to retain the basic shape of (1) and add to it in various ways, but to offer a quite different proposal, i.e. one that is a *non-modal* in character.

What might this alternative be like? I think it is fair to say that this is a currently open question in the literature. So in this section, I aim only to review three leading candidates and to note some issues with each.

10.1 Second-order properties

The first non-modal definition has been explored 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:

- (12) 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*; what is the relation between it and supervenience

physicalism? Supervenience physicalism does not entail second-order realization 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 realization physicalism entail supervenience physicalism? The usual assumption is that it does; and indeed it is natural to suppose that any proposed definition of physicalism would entail supervenience physicalism even if the reverse is not true (though see Montero 2013 for an argument to the contrary). However, 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 realization physicalism, these properties themselves will be either physical or realized by physical properties. If the first option is taken, the second-order realization 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 realization 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 realization physicalism might say that these properties supervene on physical properties. But now it hard to see the difference between the realization physicalist and the supervenient physicalist in the first place.

10.2 Subsets of causal powers

The second non-modal 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 sub-set of S^* . (We may call this notion 'sub-set realization' to distinguish it from the different notion of realization just considered.) This suggests:

- (13) Physicalism is true at a possible world w iff every property instantiated at w is either a physical property or is sub-set realized in a physical property.

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

Does sub-set 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 sub-set realize another and yet be quite different from it in nature. And this in turn suggests that sub-set realization 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 sub-set 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; at any rate, the sufficiency

objection by itself provides no reason for doubting that supervenience is necessary. Hence the status of the sub-set model remains controversial.

10.3 Grounding

The third non-modal definition 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 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:

- (14) 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 (14) 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, however, there is the further problem that the notion itself is controversial. 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. For example, if grounding physicalism, like supervenience physicalism, is intended only to capture minimal physicalism, in the sense described above (see Supervenience Physicalism as Minimal Physicalism) it may be no objection that it fails to answer these questions. Whatever is the truth about this, however, 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 unclear.

11. Understanding ‘Physical’: Introductory

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.

The condition question 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, if we say that, no two possible worlds can be physical duplicates without being duplicates simpliciter, we don't know what we've said unless we understand what it would take to be a physical duplicate, as opposed (say) to a chemical duplicate or a financial duplicate. (The point here is a quite general one: if Thales says that everything is water, or Up-to-Date-Thales says everything supervenes on water, we don't understand what he says unless he says something about what water is. The physicalist is in the same position.)

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 in the literature. 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 either is the sort of property that physical theory tells us about or else is a property which metaphysically (or logically) supervenes on 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. Similarly, if physical theory tells us about the property of being a rock — or, what is perhaps more likely, if the property of being a rock supervenes on properties which physical theory tell us about — then it too is a physical property. (The theory-based conception bears some relation to the notion of physical¹ 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 either is the sort of property required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents or else is a property which metaphysically (or logically) supervenes on the sort of property required by a complete account of the intrinsic nature of

paradigmatic physical objects and their constituents.

According to the object-based conception, for example if rocks, trees, planets and so on are paradigmatic physical objects, then the property of being a rock, tree or planet is a physical property. Similarly, if the property of having mass is required in a complete account of the intrinsic nature of physical objects and their constituents, then 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.)

It is important to note that both conceptions of the physical remain silent on the question of whether topic-neutral or functional properties should be treated as physical or not. To borrow a phrase from Jackson (1998), however, it seems best to treat these properties as onlooker properties: given any set of physical properties, one might add onlooker properties without compromising the integrity of the set. But onlooker properties should not be treated as being physical by definition.

Do these conceptions characterize the same class of properties? There are a number of different possibilities here, not of all of which we can discuss. 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). However, if this is correct, it would seem that the physical properties described by the theory conception are only a sub-class of the physical properties described by the object conception. For 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.

12. Understanding ‘Physical’: Further Issues

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 this section, I will review some main ones.

12.1 Circularity

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 as attempting to provide a reductive analysis. 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.

12.2 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). But while there is something right about this, there is also something wrong about it. What is right about it is that there is a sense in which it is rational to believe that physics is complete. After all, isn't it rational to believe that the most current science is true? But even so — and here is what is wrong about the suggestion — it is still 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.)

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.

12.3 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). Imagine the possibility of panpsychism, i.e. the possibility that all the physical objects of our acquaintance are conscious beings just as we are. 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 no matter how implausible and outlandish it sounds, 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.

12.4 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).

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 non-mental. A better way would be to require of any spelling out the notion of the physical, either the object-based account or the theory-based

account, that it respect that fact that some (uninstantiated) mental properties are non-physical.

12.5 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 the *via negativa*, 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 co-incide 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.

13. Physicalism and the Physicalist World Picture

Perhaps because of its connection to the physical sciences, physicalism is sometimes construed as an entire package of views, which contains the metaphysical thesis I have isolated for discussion as only one part. If we want a name for the entire package of views including the metaphysical claim we might call it the *Physicalist World Picture*. I will close our discussion of the interpretation question by considering the relation between physicalism (the metaphysical claim) and various other items that at least sometimes have been thought to be a part of the Physicalist World Picture.

(a) Methodological Naturalism: the idea that the mode of inquiry typical of the physical sciences will provide theoretical understanding of the world, to the extent that this sort of understanding can be achieved. Physicalism is not methodological naturalism because physicalism is a metaphysical thesis not a methodological thesis.

(b) Epistemic Optimism: the idea that the mode of understanding typical of the sciences can be used by us, i.e. by human beings, to explain the world in total, to provide a final theory of the world. Physicalism is not epistemic optimism because, since commitment to physicalism does not commit you to methodological naturalism, it clearly does not commit you to any optimism about the success of that method in the long run.

(c) Final Theory: the idea that there *is* a final and complete theory of the world, regardless of whether we can formulate it. One might

think it obvious that if physicalism is true, there is a final theory of the world. However, because of some unclarity in the notion of a theory, the issues here are not cut and dried. According to some views, something is a theory only if it is finitely stateable in a language we can understand. If that is so, clearly physicalism does not entail the idea of a final theory. On a looser conception of a theory, however, it is reasonable to say that physicalism entails that there is a final theory.

(d) **Objectivity:** the idea that the final and complete theory of world, if it exists, will not involve any essential reference to particular points of view or experiences. It is reasonable to say that physicalism entails objectivity. However, given the possibilities of non-reductive or *a posteriori* physicalism even here the issues are not settled. On those approaches, it seems possible to have irreducible points of view or experiences supervening on something physical, which compromises objectivity.

(e) **Unity of Science:** the idea that all the branches of sciences developed by us will or should be unified into a single science, usually (but not always) thought of as physics. This thesis is clearly a methodological thesis about how science ought to proceed. As we have seen, however, physicalism is a metaphysical thesis rather than a methodological thesis about how science ought to proceed. Hence it is not equivalent to the unity of science thesis.

(f) **Explanatory Reductionism:** the idea that all genuine explanations must be couched in the terms of physics, and that other explanations, while pragmatically useful, can or should be discarded as knowledge develops. Physicalism is not explanatory reductionism because, as we saw in our discussion of non-reductive physicalism, physicalism is consistent with the idea that

special sciences are quite distinct from physics. One might say that the special sciences are concerned with patterns in the physical that physicists themselves are not concerned with. For that reason the subject matter of the special sciences is distinct from the subject matter of physics.

(g) **Generality of Physics:** the idea that every particular event or process which falls under a law of the special sciences (i.e. sciences other than physics) also falls under a law of physics. In general, this view presupposes a view about laws and explanation — for example, it implies or seems to imply that special sciences have laws. But physicalism does not entail any such thesis.

(h) **Causal Closure of the Physical:** the idea that every event has a physical cause, assuming it has a cause at all. Strictly speaking, physicalists are not committed to realism about causation, so they are not committed to causal closure. (Of course, many physicalists do think that causal closure is true, as we will see below, but their position does not entail causal closure.)

(i) **Empiricism:** the idea that all knowledge (with the possible exception of conceptual knowledge) is ultimately founded on sensory or perceptual experience. Empiricism can be given a descriptive or a normative reading. On its descriptive reading, it is most likely false. Most of the information that normal humans come to deploy seems to be caused by both experience *and* inborn structure and maturation. On the normative reading, the claim is that justification is, at the end of the day, based on experience. But this epistemological thesis has nothing to do with physicalism.

(j) **Nominalism:** the idea that there are no abstract objects, i.e., entities not located in space and time, such as numbers, qualities or

propositions. If we assume that abstract objects, if they exist, exist necessarily, i.e., exist in all possible worlds, then supervenience physicalism is completely silent on the question of whether abstract objects exist. All supervenience says is that if a world is a minimal physical duplicate of the actual world, it is a duplicate simpliciter. But if abstract objects exist then they clearly exist in both the actual world and any duplicate of the actual world. What this suggests is that nominalism is a distinct issue from physicalism (Schiffer 1987, Stoljar 1996).

(k) **Atheism:** the idea that there is no God as traditionally conceived. In the 17th and 18th century, physicalism (or materialism, as it was then known) was widely but not universally viewed as inconsistent with belief in God (Yolton 1983). Nowadays, this issue is somewhat less discussed. Nevertheless, as we noted previously, *if* God is thought of as essentially non-physical, then Atheism *does* seem to be a consequence of physicalism, at least on some interpretations of the background modal notions.

14. The Case Against Physicalism I: Qualia and Consciousness

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.

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. (There are also a number of other arguments in this area — for a very good recent discussion, see Chalmers 1996). 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 posteriori* physicalism is that (3) — i.e. the claim that *S* entails *S** — is *a posteriori*. Since (3) 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 (3). On the other hand, the mere fact that Mary has not had (and cannot have) the experience to know (3) does not remove the possibility that (3) is true. Hence *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 (3) is physical or not: if (3) is physical, then the response attacks P1. But if (3) 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.) 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 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 (cf. Chalmers 1996, Lockwood 1992, and Stoljar 2000, 2001.)

15. The Case Against Physicalism II: 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. 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).

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*.

16. The Case Against Physicalism III: 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 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 completely unclear where philosophical 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 the details of the project — that would not distinguish them from working scientists. Rather they are concerned with what the potential limits of the project are.

This is a theme which has been pursued by Thomas Nagel (1983) and allied work by 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.

17. The Case for Physicalism

Having considered one side of the truth question, I will now turn 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 claim that everything supervenes on the physical.

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. 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 it 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, once it is appreciated what physicalism is — and, more important, what it is not — 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.

18. Further Questions

That completes our discussion of physicalism. It is sketchy in certain places, and this suggests that there is much further work to be done before we arrive at a final assessment of the doctrine of physicalism and the role that it plays in contemporary thought. In the future, however, it would seem to me to be a good idea to think at least about the following:

(a) The contrast between modal and non-modal definitions of physicalism. While supervenience, and more generally modal, definitions of physicalism are attractive, they do raise problems of a difficult nature—such as the blockers problem and the sufficiency problem. At the time of writing it is unclear how these difficulties will be resolved.

(b) The relation between, and the adequacy of, the theory-conception and the object-conception of the physical. It seems clear that our thinking about the physical is anchored in part in the ordinary idea of a physical object and in part in the idea of physics. But it seems equally clear that these two ideas can and do come apart. What are the consequences of this for the notion of physicalism?

(c) The role of physicalism in philosophy of mind. As we have seen for both Nagel and Williams questions about physicalism are really questions about objectivity in disguise. This raises the more general issue of whether arguments and theses in philosophy of mind (and other areas of philosophy) that are formulated in terms of physicalism are necessarily formulated in that way. Perhaps physicalism here is simply getting in the way of a proper understanding of how these problems work?

Bibliography

- Armstrong, D., 1968, *A Materialist Theory of the Mind*, Lond: Routledge.
Bacon, J., 1990, 'Van Cleve Versus Closure', *Philosophical Studies*, 58:
239–242.

- Bennett, K., 2003, 'Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it', *Noûs*, 37(3): 471–497.
- Berkeley, G., 1710, *Principles of Human Knowledge*, London: Penguin, 2004.
- 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.
- , 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.
- Chomsky, N., 1994b, ‘Noam Chomsky’, in S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Oxford: Blackwell.
- , 1995, ‘Language and Nature’, *Mind*, 104(413): 1–61.
- , 2000, *New Horizons in the Study of Language and Mind*, Cambridge: Cambridge University Press.
- Correia, F and Schneider, B (eds.), 2012, *Metaphysical Grounding: Understanding the Structure of Reality*, Cambridge: Cambridge University Press.
- Crane, T. and Mellor, D.H., 1990, ‘There is no Question of Physicalism’, *Mind*, 99: 185–206.
- Daly, C. 1997, ‘What are Physical Properties?’, *Pacific Philosophical Quarterly*, 79(3): 196–217.
- Davidson, D. 1970 . ‘Mental Events’, in D. Davidson, *Essays on Actions and Events*, Oxford: Oxford University Press, 207–223.
- Descartes, R., 1641, *Meditations on First Philosophy*, in Cottingham *et al.* (eds.), *The Philosophical Writings of Rene Descartes*, Cambridge: Cambridge University Press, 1985.
- Dijksterhuis, E.J., 1961, *The Mechanization of the World-Picture*, Oxford: Clarendon.
- Dowell, J.L., 2006a, ‘Formulating the Thesis of Physicalism’, *Philosophical Studies*, 131(1): 1–23.
- , 2006b, ‘Physical: Empirical not Metaphysical’, *Philosophical Studies*, 131(1): 25–60.
- Feigl, H., 1967, ‘The “Mental” and the “Physical”’, Minneapolis: University of Minnesota Press, 370–497.
- Feinberg, G., 1966, ‘Physics and the Thales Problem’, *Journal of Philosophy*, 63: 5–16.
- Field, H., 1972, ‘Tarski’s Theory of Truth’, *Journal of Philosophy*, 69:

347–75.

- , 1992, ‘Physicalism’, in J. Earman (ed.), *Inference, Explanation and Other Frustrations*, Berkeley: University of California Press, 271–292.
- Fine, K., 2001, ‘The Question of Realism’, *Philosophers' Imprint*, 1: 1–30.
- Fodor, J.A., 1974, ‘Special Sciences: Or, The Disunity of Science as a Working Hypothesis’, reprinted in J. Fodor, *Representations*, Cambridge, MA: MIT Press, 1981.
- , 1992, *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press
- Foster, J., 1982, *The Case for Idealism*, London: Routledge
- , 1991, *The Immaterial Self: A Defence of the Cartesian Dualist Conception of Mind*, London: Routledge.
- Gendler, T. and Hawthorne, J., 2004, *Conceivability and Possibility*, Oxford: Oxford University Press
- Gillet, C. and Loewer, B., 2001, *Physicalism and Its Discontents*, Cambridge: Cambridge University Press
- Haugeland, J., 1983, ‘Weak Supervenience’, *American Philosophical Quarterly*, 19: 93–103.
- Hawthorne, J., 2002, ‘Blocking Definitions of Materialism’, *Philosophical Studies*, 110(2): 103–113
- Hellman, C., 1985, ‘Determination and Logical Truth’, *The Journal of Philosophy*, 82(11): 607–616.
- Hempel, C. 1949, ‘The Logical Analysis of Psychology’, in H. Feigl and W. Sellars (eds.), *Readings in Philosophical Analysis*, New York: Appleton-Century-Crofts, 1949, pp. 373–384; reprinted in N. Block (ed.), *Readings in the Philosophy of Psychology*, Volume 1, Cambridge: Harvard University Press, 1980, pp. 14–23.
- , 1969, ‘Reduction: Ontological and Linguistic Facets’, in S. Morgenbesser, et al. (eds.), *Essays in Honor of Ernest Nagel*, New York: St Martin's Press.

- , 1980, 'Comments on Goodman's *Ways of Worldmaking*', *Synthese*, 45: 139–199.
- Hohwy, J., 1998, *Meaning as Use*, Ph.D Dissertation, Australian National University.
- Horgan, T., 1983, 'Supervenience and Microphysics', *Pacific Philosophical Quarterly*, 63: 29–43.
- , 1993, 'From Supervenience to Superdupervenience: Meeting the Demands of a Material World', *Mind*, 102(408): 555–586.
- Horwich, P., 2000, *Meaning*, Oxford: Oxford University Press.
- Hyslop, A., 1999, 'Methodological Epiphenomenalism', *Australasian Journal of Philosophy*, 78(1): 61–70.
- Jackson, F., 1982, 'Epiphenomenal Qualia', *Philosophical Quarterly*, 32: 127–36.
- , 1986, 'What Mary Didn't Know', *Journal of Philosophy*, 83: 291–5.
- , 1998, *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, Oxford: Clarendon.
- , 1993, 'Armchair Metaphysics', in J. Hawthorne and M. Michael (eds.), *Philosophy in Mind*, Amsterdam: Kluwer.
- Jackson, F., and Pettit, P., 1992, 'In Defense of Explanatory Ecumenism', *Economics and Philosophy*, 8: 1–21.
- Kim, J., 1993, *Mind and Supervenience*, Cambridge: Cambridge University Press.
- , 1998, *Mind in a Physical World*, Cambridge: Cambridge University Press.
- , 2005, *Physicalism, Or Something Near Enough*, Princeton: Princeton University Press.
- Kripke, S., 1980, *Naming and Necessity*, Cambridge, MA: Harvard University Press.
- , 1982, *Wittgenstein on Rules and Private Language: An Elementary Exposition*, Oxford: Basil Blackwell.
- Lange, F., 1865, , *A History of Materialism*, London: Routledge, 1925.

- Latham, N., 2001, 'Substance Physicalism', in C. Gillett and B. Loewer (eds.), *Physicalism and Its Discontents*, Cambridge: Cambridge University Press, pp. 152–171.
- Leuenberger, S., 2008, 'Ceteris Absentibus Physicalism', in D. Zimmerman (ed.), *Oxford Studies in Metaphysics*, Oxford: Oxford University Press, pp. 145–170.
- Lewis, D., 1970, 'How to Define Theoretical Terms', *Journal of Philosophy*, 67: 427–46.
- , 1983, 'New Work for a Theory of Universals', *Australasian Journal of Philosophy*, 61(4): 343–377.
- , 1986, *On the Plurality of Worlds*, Oxford: Blackwell.
- , 1994, 'Reduction of Mind', in S. Guttenplan (ed), *A Companion to the Philosophy of Mind*, Oxford: Blackwell, pp. 412–431.
- , 1997, 'Naming the Colours', *Australasian Journal of Philosophy*, 75: 325–342.
- Loar, B., 1997, 'Phenomenal States', in N. Block, *et al.* (eds.), *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press.
- Lockwood, M., 1989, *Mind, Brain and Quantum*, Oxford: Basil Blackwell.
- Ludlow, P., Nagasawa, Y. and Stoljar, D., 2004, *There's something about Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, Cambridge: MIT Press.
- Lycan, W., 1996, *Consciousness and Experience*, Cambridge, MA: MIT Press.
- MacLaughlin, B., 1992, 'The Rise and Fall of British Emergentism', in A. Beckermann *et al.* (eds.), *Emergence or Reduction?*, Berlin: De Gruyter.
- Maudlin, T., 1996, 'On the unification of physics', *Journal of Philosophy*, 93: 129–144.
- Meehl, P.E. and Sellars, W.S., 1956, 'The Concept of Emergence', in H.

- Feigl and M. Scriven (eds.), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis* (Minnesota Studies in the Philosophy of Science: Volume 1), Minneapolis: University of Minnesota Press, 239–252.
- Melnyk A., 1997, “How To Keep The ‘Physical’ in Physicalism”, *Journal of Philosophy*, 94: 622–637.
- , 2003, *A Physicalist Manifesto: Thoroughly Modern Materialism*, Cambridge: Cambridge University Press
- Montero, B., 2013, “Must Physicalism imply the supervenience of the mental on the physical?”, *Journal of Philosophy*, 110(2): 93–110.
- Montero, B. and Papineau, D., 2005, “A defense of the Via Negativa Argument for Physicalism”, *Analysis*, 65(3): 233–237.
- Nagel, E., 1961, *The Structure of Science*, New York: Harcourt, Brace and World.
- Nagel, T., 1974, ‘What is it like to be a bat’, *Philosophical Review*, 4: 435–50.
- , 1983, *The View from Nowhere*, New York: Oxford.
- Nemerow, L., 1988, ‘Physicalism and the Cognitive Role of Acquaintance’, in W. Lycan (ed.), *Mind and Cognition*, Oxford: Blackwell.
- Neurath, O., 1931, ‘Physicalism: The Philosophy of the Vienna Circle’, in R.S. Cohen, and M. Neurath (eds.), *Philosophical Papers 1913–1946*, Dordrecht: D. Reidel Publishing Company, 1983, pp. 48–51.
- Ney, A., 2008, ‘Physicalism as an Attitude’, *Philosophical Studies*, 138: 1–15.
- Papineau, D., 1996, *Philosophical Naturalism*, Oxford: Blackwell.
- Paull, C., and Sider, T., 1992, ‘In Defense of Global Supervenience’, *Philosophical and Phenomenological Research*, 52: 833–854.
- Poland, J., 1994, *Physicalism: The Philosophical Foundations*, Oxford: Clarendon.
- Putnam, H., 1975, ‘Philosophy and our mental life’, in H. Putnam, *Mind*,

- Language and Reality: Philosophical Papers*, Volume 2, Cambridge: Cambridge University Press.
- Rosen, G., 2010, 'Metaphysical Dependence: Grounding and Reduction', in B. Hale and A. Hoffman (eds.), *Modality: Metaphysics, Logic and Epistemology*, Oxford: Oxford University Press, pp. 109–36
- Russell, B., 1917, 'On the Notion of Cause', in B. Russell, *Mysticism and Logic*, London: Penguin, 1963.
- , 1927, *The Analysis of Matter*, London: Kegan Paul.
- Ryle, G., 1949, *The Concept of Mind*, London: Routledge.
- Schaffer, J., 2009k, 'On What Grounds What', in D. Chalmers, D. Manley, R. Wasserman (eds), *Metametaphysics*, New York: Oxford University Press. pp. 247–383
- Schiffer, S., 1987, *Remnants of Meaning*, Cambridge, MA: MIT Press.
- Searle, J., 1992, *The Rediscovery of the Mind*, Cambridge, MA: MIT Press.
- , 1998, *Mind, Language and Society: Philosophy in the Real World*, New York: Basic Books.
- Shoemaker, S., 1994, 'Phenomenal Character', *Noûs*, 28: 21–38
- , 2007, *Physical Realization*, Oxford: Oxford University Press.
- Smart, J.J.C., 1959, 'Sensations and Brain Processes', reprinted in D. Rosenthal (ed.), *Materialism and the Mind-Body Problem*, Indianapolis: Hackett, 1987.
- , 1978, 'The Content of Physicalism', *Philosophical Quarterly*, 28: 239–41.
- Smith, M., and Stoljar, D., 1998, 'Global Response-Dependence and Noumenal Realism', *The Monist*, 81(1): 85–111.
- Stanley, J., and Williamson, T., 2001, 'Knowing How'. *Journal of Philosophy*, 98: 411–444
- Stalnaker, R., 1996, 'Varieties of Supervenience', *Philosophical Perspectives*, 10: 221–241.
- Steward, H., 1996, *The Ontology of Mind*, Oxford: Clarendon.

- Stoljar, D., 1996, 'Nominalism and Intentionality', *Noûs*, 30(2): 261–281.
- , 2000, 'Physicalism and the Necessary *A Posteriori*', *Journal of Philosophy*, 97(1): 33–54.
- , 2001, 'Two Conceptions of the Physical', *Philosophy and Phenomenological Research*, 62: 253–281.
- , 2001a, 'The Conceivability Argument and Two Conceptions of the Physical', *Philosophical Perspectives*, 15: 393–413.
- , 2010, *Physicalism*, New York: Routledge.
- Stroud, B., 1986, 'The Physical World', *Proceedings of the Aristotelian Society*, 87: 263–277.
- Van Cleve, J., 1990, 'Supervenience and Closure', *Philosophical Studies*, 58: 225–283.
- Van Fraassen, B., 2002, *The Empirical Stance*, New Haven: Yale University Press.
- Williams, B., 1985, *Ethics and the Limits of Philosophy*, London: Fontana.
- Wilson, J., 1999, 'How Superduper does a Physicalist Supervenience need to be?', *Philosophical Quarterly*, 49: 33–52.
- , 2005, 'Supervenience-based formulations of Physicalism', *Nous*, 39:3: 426–459.
- , 2006, 'On Characterizing the Physical', *Philosophical Studies*, 131: 61–99.
- , 2010, 'What is Hume's Dictum and Why Believe it', *Philosophy and Phenomenological Research*, 80: 595–637.
- , 2011, 'Non-reductive Realization and the Powers-based sub-set strategy', *The Monist*, 94: 121–154.
- , 2014, 'No work for a Theory of Grounding', *Inquiry*, 57: 1–45.
- Yablo, S., 1992, 'Mental Causation', *The Philosophical Review*, 101: 245–280.
- , 1999, 'Concepts and consciousness', *Philosophy and Phenomenological Research*, 59: 455–464.
- Yolton, R., 1983, *Thinking Matter*, Minneapolis: University of Minnesota

Press.

Academic Tools

-  How to cite this entry.
-  Preview the PDF version of this entry at the Friends of the SEP Society.
-  Look up this entry topic at the Indiana Philosophy Ontology Project (InPhO).
-  Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

- Physicalism, bibliography at PhilPapers.org.
- Philosophy of Mind, General Works, bibliography at PhilPapers.org.

Related Entries

behaviorism | color | Davidson, Donald | epiphenomenalism | multiple realizability | qualia | supervenience

Acknowledgments

The author would like to thank Hossein Ameri, Tim Bayne, Rich Cameron, Brian Garrett, Robert Pasnau, Stewart Saunders, Jessica Wilson, and particularly David Chalmers for their help in constructing this entry. In addition, the author and editors would like to thank two readers, Joshua R. Stern and Greg Stokley, for discovering numerous typographical errors on an earlier version. Their volunteer efforts were entirely unsolicited and very much appreciated.

Copyright © 2017 by the author

DANIEL STOLJAR

Daniel Stoljar