9 Physicalism and the A Priori

Frank Jackson

1 A Bit of History

Jack Smart and David Armstrong told me about physicalism and the identity theory of mind back in the 1960s. The version they espoused held inter alia that the physical way things are a priori entails the mental way things are, though they didn't say it quite that way. They talked instead of the way certain analyses of mental states in topic-neutral or functionalist or central-state terms show how one could in principle deduce mental nature from enough information of the right kind about physical nature and especially neural nature.[1] As we say it these days, they espoused a priori, or *A*-type physicalism, and not a posteriori, or *B*-type physicalism.[2]

They convinced me that a priori physicalism is the physicalism of choice; indeed, that a physicalist pretty much has to be an a priori physicalist. However, many physicalists these days are avowedly a posteriori physicalists, often being inspired by the example of the relation between water and H_2O. Just as, they say, the distribution of H_2O necessitates without a priori determining the distribution of water, so the physical way our world is necessitates without a priori determining the mental way it is.[3] In the past I have said a number of things in favor of a priori physicalism and in favor of the idea that if one is to be a physicalist, one should be an a priori physicalist, things that Ned Block has contested for one reason or another.[4] I revisit one of our areas of disagreement in this chapter. But my main interest is, not to revisit past disagreements, but to urge that the issue between a priori and a posteriori physicalism is often misunderstood. It is not—or should not be—thought of as a dispute between two versions of a thesis about the mind as such, in the sense of a thesis about the metaphysics of mind. As far as the *metaphysics* goes, a physicalist has to be an a priori physicalist. Being an a posteriori physicalist is not a live option—or so I argue. The real issues concern the words we use to talk about mental states and the nature of mental concepts. At a

I am indebted to many, many discussions and to detailed comments from Adam Pautz and Daniel Stoljar. Almost every page bears the marks of these discussions and comments.

number of points, I note the bearing of our discussion on the challenge that the knowledge and zombie arguments pose for physicalism.

First, I need to say something about how I understand physicalism.

2 Physicalism as the Denial of Tailor-Made Mental Properties

A gauge carries information about the amount of gas left in a tank by virtue of the way its pointer covaries with the amount of gas left in the tank. Exactly how to spell this out is controversial. How reliable does the covariation have to be? What about chance and probabilistic connections? What about cases in which there is no covariation but the objects in question were *designed* to covary? And so on. Despite all this controversy, one thing seems clear: carrying information, that property, is not tailor-made for the job. Whatever property it is, we can pick it out without using the very notion of information.[5] *A* gets to carry information about *B* by virtue of facts that we can specify without using the term "information" or the concept *information* itself.

We will think of physicalism (about the mind) in like manner. The key difference between physicalism and dualism is that the latter affirms and the former denies that mental properties are tailor made. In this sense, mental properties are not fundamental. We can say this in terms of the "what God would have to do" way of thinking of the divide between physicalism and dualism. Physicalism says that God could make something with a mind, drawing solely on properties we can specify without using the very notions of, say, pain or belief or looking red. It is enough to assemble in the right way ingredients whose properties and modes of assembly earn their keep in contexts outside psychology. By contrast, dualism says that God would have to use a property or a mode of assembly that can be specified only in terms of the notions of pain, consciousness, belief, looking red, and so on.

Why believe in physicalism as just specified? That's outside my brief, but allow me a quick observation. It is surely very plausible that the process that starts with conception and ends up with creatures like us is precisely a process of assembling, in the right way, ingredients that earn their keep outside psychology as such, using modes of composition that earn their keep outside psychology, and we know that this process delivers creatures with minds.

Why did I specify physicalism as the denial of tailor-made mental properties? Why not simply say that physicalism affirms that mental properties are one and all properties that *supervene* on physical properties, adding a remark or two about physical properties being those that play a role in physical theory? This, after all, is what is often done and is in part responsible for the name "physicalism." I have two reasons. One is a desire to avoid needless controversy. As is well known, it is surprisingly difficult to say what playing a role in physical theory comes to in a way that allows physicalism to meet, simultaneously, two constraints: that of being a theory that has a good chance of

being true and that of being a theory with enough bite to be worth supporting. What's important for what follows is not physicalism thought of as tied to some potentially troublesome notion of playing a role in physical theory but physicalism thought of as the doctrine that minds and their properties are constructions out of nonminds and nonmental properties. The second reason relates to an issue that regularly comes up when physicalism is specified in terms of supervenience. Many agree that the supervenience of mental properties on physical properties is a *necessary* condition for the truth of physicalism (while disagreeing about important details); the big controversy is over whether it is a *sufficient* condition for the truth of physicalism. Some theists hold that God exists of necessity. It follows that God's existence supervenes on everything, including the physical. But surely this doesn't mean that theists are physicalists despite their protestations to the contrary. Or consider the version of dual-attribute dualism that holds that the physical metaphysically necessitates the instantiation of special properties that constitute our mental natures (or maybe just the phenomenal side of our mental natures). Supporters of this position believe in the supervenience of the mental on the physical but are dualists by their own explicit avowal. Of course, if the versions of theism and dualism just mentioned are necessarily false or a priori false, they are no threat to the claim that there is a suitable formulation of the supervenience of the mental on the physical that is true if and only if physicalism is true. All the same, an important part of the content of physicalism is not addressed by supervenience formulations of physicalism. It is for this reason that one species of antireductionist theory of mind does not seem to me to be a version of physicalism. I mean here the species that insists that there are no interesting identifications of mental properties to be had, let alone identifications of mental properties with functional or neurological properties or anything like that, but that the mental supervenes on the physical.

We will, therefore, think in what follows of physicalism as the thesis that mental properties are identical with properties we can specify without using mental concepts—this is not, of course, to be read as saying that mental properties are identical with nonmental properties; that would be a violation of Leibniz's law—and will use the term "physical properties" to pick out this class of properties.[6] And this is, in practice, what physicalism typically comes to in the debates between physicalists and dualists. The physicalists affirm and the dualists deny that the kind pain, as it might be, is a certain neurological kind, or is a disjunction of same, or is a certain functional kind, when it is clear that although these kinds are typically specified in highly nonspecific terms (lots of hand waving goes on), the terms in question are one and all ones that earn their keep outside psychology. Neurological and functional kinds, for example, aren't the preserve of psychology alone. Having said this, in what follows I focus the discussion by framing it in terms of mental properties thought of as neurological ones or as functional ones. I don't think I beg any key questions by doing this.[7] Also, the focus is on certain sorts of mental properties, those often tagged "phenomenal" or "experiential,"

those often said to be those for which there is something it is like to exemplify them. However, I note near the end that the "something it is like" way of picking out the class of mental properties in question is potentially misleading in an important respect.

I can now state a determination thesis common to both a priori and a posteriori physicalism.

3 The Determination Physicalists (Nearly All) Agree On

Physicalists disagree about a lot, but one idea unites them: in some strong sense, the mental is exhausted by the physical: given enough physical nature (including the physical laws), mental nature is determined without remainder. This follows from the core idea that the mental way we are is not an addition to the physical way we are and, indeed, is implicit in the talk above of physicalism denying that the mental is tailor made. Once we have said—or granted—this much, we have an easy way to describe the difference between a priori and a posteriori physicalism. The first says that the determination from physical to mental, however we spell it out, taking account of all the niceties, is like the determination of the shape of an array of points by the location of those points (and maybe the curvature of the space): it is a priori as well as necessary. The second says that it is like the determination of the distribution of water (gold) by the distribution of H_2O (stuff with atomic number 79): it is necessary but a posteriori.

Here we would seem to have a distinction between two theses in the metaphysics of mind to do with relations between physical and mental properties, with the second thesis explained by reference to a widely accepted thesis about the *nature* of water—as the distribution of H_2O is to the distribution of water, so the distribution of the physical is to the distribution of the mental. But it is (of course) properties and items with properties that get to be *distributed* in one way or another. The same goes for giving the key idea using the phrase "the way things are." Thus, we often characterize a posteriori physicalists as holding that the physical way things are determines without entailing (in the a priori deducibility sense) the mental way things are. But, of course, "the ways things are" manner of speaking is a linguistic device for speaking of properties and things with the properties.

However, as I detail in the section to come, *read as a thesis in the metaphysics of mind*, a posteriori physicalism cannot be true. As we will observe, a certain kind of dualist can affirm that physical properties necessitate without a priori entailing mental properties, but it is not a thesis that physicalists can embrace.

4 Against the Metaphysical Reading of A Posteriori Physicalism

Everyone (almost) acknowledges that a signal contribution of Saul Kripke (1980) was to convince us that there are necessary a posteriori truths (and contingent a priori ones; more on that later). But some insist that what we learn about is a feature of sentences,

not of propositions; or more exactly, that the convincing examples he, and those following him, drew to our attention concern sentences. Perhaps there are examples elsewhere of propositions that are necessarily a posteriori true.[8] Others insist that Kripke's message cuts much deeper and that the "sentence brigade" just don't get it. However this debate should be resolved, we can all agree (as many have noted) that those who insist that we learned something about propositions need to be careful to avoid saying things that implicitly commit them to the following inconsistent triad:

The proposition *that Eric Blair = George Orwell* is a posteriori true.

The proposition *that Eric Blair = Eric Blair* is a priori true.

The proposition *that Eric Blair = George Orwell* is the same proposition as that of the proposition *that Eric Blair = Eric Blair*.

This triad violates Leibniz's law. A similar point applies to the corresponding beliefs, of course. We should, for example, avoid saying anything that commits us to holding that the belief that Eric Blair is George Orwell is the same belief (in the sense of *what's* believed) as the belief that Eric Blair is Eric Blair but that the first belief is a posteriori true whereas the second is a priori true.

The issue here is not one about intentionality or hyperintentionality or anything in that neck of the woods. Many would want to say of the following two *sentences*

I believed in 1960 that Eric Blair was Eric Blair

and

I did not believe in 1960 that Eric Blair was George Orwell

that the first might be true when the second is false (perhaps I did not learn that Blair was Orwell until 1961) at the same time as holding that Blair's being Orwell and Blair's being Blair not only have the same truth value but do so necessarily.[9] I am not taking a position on what to say about examples like these except to say that the one thing one should not do is explain the (alleged) phenomenon by saying that there is a proposition that has both the property of being believed by me in 1960 and not being believed by me in 1960.

The same goes for the (pretty standard) claim that water's not being H_2O is metaphysically impossible while being conceptually possible, whereas being H_2O without being H_2O is both metaphysically and conceptually impossible. This has to be spelled out in a way that makes it clear that one is *not* holding at once the following:

Being water is the same property as being H_2O.

Being H_2O necessitates without a priori necessitating being water.

Being H_2O a priori necessitates being H_2O.

No doubt this can be done. Whether it can be done without joining the camp of those who just don't get it is another question and one for another time.[10]

In discussion, two objections come up at this point. First, some object that relations like necessitating and a priori necessitating between properties are relative to the way we pick out the properties. They say that if Fred is thinking of the property of being taller than three people and George of the property of being taller than at least two people, then although being taller than three people a priori necessitates being taller than at least two people, it is not true that the property Fred is thinking of a priori necessitates the property George is thinking of. But, of course, it *is* true. What isn't true is something else altogether. Being the property Fred is thinking of does not a priori necessitate being the property George is thinking of, but being the property Fred is thinking of is not being taller than at least three people, even in the case where that's the property he is thinking of. It is rather the property of being whatever property George is thinking of.[11]

The second objection is that it is a mistake to think of a priori necessitation as a relation that may or may not hold between properties: maybe necessitation is, but the category of the a priori pertains to sentences or predicates or concepts and not in any case to properties literally speaking. There is no *de re* a priori necessitation, as it might be put.

I think this has things the wrong way around. Surely, among the central questions in metaphysics are the following: Is there *de re* vagueness? Is unrestricted composition true? Can an object be two different colors at once? Is freedom compatible with determinism? Does having a shape imply being extended? None of these questions concern words or concepts as such. They are properly speaking questions in *metaphysics*. Many hold that the answer to one or another of these questions is an a priori matter. They may be wrong, but not, surely, because the questions aren't about words and are about things and properties. If that was the right thing to say, we could say a priori that there are no a priori issues in metaphysics proper.

Why, then, might one hold that the sentence "Anything with a shape has an extension," say, is a priori? Because one held that understanding the words in the sentence involved a grasp of the properties designated by them, *and* one held that the properties in question were such that being shaped a priori necessitated being extended. That is to say, what is fundamental are the relations between the properties and more generally the metaphysics; the status, a priori or otherwise, of the sentence comes second, or so it seems to me.

I can now make my key point very quickly. One thing the a posteriori physicalist cannot maintain is that physical properties can at most necessitate mental properties; they never a priori necessitate them. Mental properties *are* physical properties according to both a priori and a posteriori physicalists, and every property a priori necessitates itself. Physicalists have to allow that some physical properties a priori necessitate mental properties. They may differ over which physical properties a priori necessitate which mental properties—is it the functional ones, or the neurological ones, or the biological

properties with such and such an evolutionary history, and so on, or is it some combination of them?

What we need to do next is obvious. We need to think of a posteriori physicalism as a doctrine about something other than relations between properties. In the sections to come, we look at the doctrine thought of as one about sentences and the doctrine thought of as one about concepts. There are other possibilities, but those two will keep us busy enough. But first I note an unhappy consequence of the conclusion that the a posteriori physicalism cannot be thought of as a doctrine in metaphysics or, I should say, an unhappy consequence for those who wish to use a posteriori physicalism to blunt the force of the zombie and knowledge arguments. Dualists won't be unhappy.

5 What, Exactly, Are the Zombie and Knowledge Arguments *About*?

When something looks red to us and we are aware of this fact, we are in a highly distinctive kind of state, and we know that the state differs from the state we are in when something looks green or when we are in pain and are aware of this. If that wasn't true, we would not know what to expect when we hear words like "In a few minutes you will see a red sunset," "After the rain everything will look green," or "This is going to hurt." But we do know what to expect—that is, what kind of state to expect. We, that is, do not directly refer to the state in a sense of "directly" tied to referring to it minus any grasp of the *kind* of state it is. Of course, it may well be true that we refer to the state but not via properties the state has. But that is true of our thought that something is square. When we think that something is square, we do not think that it has the property that is thus and so, where thus and so is a property *of* being square *distinct* from being square itself. We think that it is *square*.

The challenge of the zombie and knowledge arguments is that it very much seems that this property—the one that we know we instantiate when something looks red to us and are aware of the fact—is quite different from the kinds of properties physicalists talk about in their accounts of the mind. This is why it seems possible to have all the physical properties without having the "something looks red, and I'm aware of the fact" property (the zombie case) and why it seems that Mary did not and could not know about the property when she was in the black and white room (the knowledge argument case). There may be some deep error in both these arguments; indeed, as a physicalist I am committed to saying just this. The property is, I have to say, a physical property, despite the intuitive appeal of the intuitions that lie behind the zombie and knowledge arguments. But my point here is that both arguments concern intuitions about (instantiated) properties; they do not concern intuitions about sentences or concepts.[12] Once we start to think of a posteriori physicalism as a thesis about sentences or about concepts, we invite the thought that, whatever may be the interest of a posteriori physicalism per se, it may well not even engage with the knowledge argument or the

zombie argument. And, in fact, some comments by those who would classify themselves as a posteriori physicalists invite the thought that they are avoiding engagement with the essential thought behind, for example, the zombie argument.

Take the following passage from a paper by David Papineau directed at the zombie argument: "It is conceivable that your physical properties could be duplicated without duplicating your conscious properties, but it does not follow that the latter are ontologically distinct from the former" (2011, 178). But what, exactly, is the situation we are being told is conceivable? If, as we physicalists say, conscious properties *are* physical properties, the situation would seem to be one with duplication in physical properties without duplication in physical properties, and that is not conceivable (in the relevant respect, we can of course think about it to the extent of declaring that it is a priori impossible). To the extent that the zombie objection to physicalism is about conscious properties *as such*, as opposed to, say, ways of referring to them or conceptualizing them, what we have here does not, it seems to me, come to grips with it.

Presentations of the knowledge argument can obscure that it is an argument resting on an intuition about properties. Mary's situation when in the black and white room is sometimes described as knowing all there is to know that can be made available using sentences containing words drawn from a certain vocabulary, in a way that suggests (or might suggest) that what is available to her is "information under a description."[13] But what is available to her in the black and white room is information about the distribution of physical properties. Much of it is delivered by words, the words drawn from that vocabulary (but she has, recall, access to black and white pictures and videos, as well as to passages of prose), but what gets delivered concerns properties and the things that have those properties. (When someone uses the word "square" to tell you how something is, what they tell you is not anything about the word as such.) It would be a misunderstanding to think that it was to the point whether the books Mary read were in English or in Greek.

It is now time to look at a posteriori physicalism thought of as a doctrine about sentences (and words) and as a doctrine about concepts.

6 As a Doctrine about Words and Sentences

I have already noted, in effect, that we should not hold, at one and the same time, the following:

1. Being H_2O = being water.
2. Being water a priori necessitates being water.
3. Being H_2O does not a priori necessitate being water.

Of course, consistency can be restored by replacing (1) by

1*. Being $H_2O \neq$ being water,

but then it is not clear that (2) would remain true.

However, these worries do not touch someone who wants to affirm certain claims about *sentences* containing the words "H_2O" and "water" of a kind that often come up when the H_2O-water case is discussed. Consider, for instance,

4. H_2O covers 70 percent of the earth

and

5. Water covers 70 percent of the earth.

One might well maintain that it is a necessary truth that if (4) is true, then (5) is true, but it is not a priori true (the conditional with (4) as antecedent and (5) as consequent is not a priori).

This might suggest we should think of a posteriori physicalism as the doctrine that although there are sentences couched in physical vocabulary that entail a sentence couched in mental vocabulary, like "Something looks red to Mary now and she is aware of the fact," there aren't ones that a priori entail that sentence.[14]

There is a lot to say about this idea. Here I restrict myself to saying why I do not think the water-H_2O example is a good model for a posteriori physicalism, understood as a thesis about the absence of a priori entailments from sentences couched in physical terms to sentences couched in mental (and especially phenomenal) terms. Does this matter much? A theory can be correct even if the example that inspired it falls over. I think it matters when the example played such a big role in inspiring it and is still regularly used to explain the key idea behind the theory. Be this as it may, let us look at the argument to the conclusion that the water-H_2O example is not a good model.

I start with something we learn from Kripke (1980), or so it seems to me. We learn that there are contingent a priori sentences. And I think—I say it roughly at first and talk about qualifications shortly—that

6. Water is the watery stuff

is an example.

Here is why one might think (6) is contingent a priori. We have known for a very long time that there is a natural kind that is potable, fills the oceans, is liquid at room temperature, and all that. We (English speakers) used the word "water" for that kind. We said things like, "There is water in that glass," and in saying things like that, we made a claim about how things are, and it was, of course, a claim we were entitled to hold true on occasion. This means that it wasn't a claim to the effect that there is H_2O in the glass. We were not, back then, entitled to hold that there is H_2O in the glass. That indeed is why it was necessary to carry out all those experiments. What we were entitled to believe back then is that what fills the glass is the natural kind that is potable, fills the oceans,

is liquid at room temperature, and all that. This strongly suggests that "water" is a name for the watery stuff, whereas "watery" is a convenient shorthand for the potable kind that falls from the sky and all that. Thus, (6) is a priori; it reflects a semantic decision about how to use a word. It is not a substantive claim about how things are. However, because it is a name, it is rigid. It refers to the same kind in every possible world. This means that it refers to H_2O in every possible world. It follows that (6) is contingent, for there are possible worlds where H_2O is not the watery stuff.[15]

Suppose (6) is indeed a priori. Then the inference from (4) and (7) to (5) is a priori the following:

4. H_2O covers 70 percent of the earth.
7. H_2O is the watery stuff.
5. Water covers 70 percent of the earth.

The upshot is that sentences couched in terms of "H_2O" *do* lead a priori to a sentence about how things are couched in terms of "water."

How might one reply to this argument? First, one might hold that it confuses a semantic question with a metasemantic question. One question is what a word means; another is how it gets to mean what it means.[16] The externalist message of the Twin Earth thought experiment (Putnam 1975) is that we should think of the meaning of "water" in terms of a function that goes from a context—in this case, whatever it is that is the watery stuff—to the kind. On Earth, this function goes to H_2O. (On Twin Earth it goes to XYZ.) The idea that "water" means anything like a definite description along the lines of "the watery stuff," be it rigidified or not, fails to take on board the clear message of the Twin Earth thought experiment.

But what—exactly—is the clear message? Here is how one pretty standard version of the thought experiment goes.[17] Fred lives on Earth, next to a lake that periodically dries up. On Monday, he looks out the window and says, "There is plenty of water out there." While asleep, he is transported to Twin Earth. On Tuesday he wakes up in a bedroom that is indistinguishable from his own and in an environment that is, as far as he is concerned, indistinguishable from the one he inhabited on Monday. He has no idea that he is no longer on Earth. He looks out the window and says, "There is plenty of water out there." What he says on Tuesday is clearly false, runs the argument, but there is plenty of watery stuff in the (different) lake he is now looking at.

But where does the judgment that what he says is false come from? He might, after all, be using the word "water" in a way that means that the sentence "There is plenty of water out there" in his mouth on Tuesday is true. As many have noted, the message of Twin Earth cannot be that Fred is unable to use "water" in a way that means it refers to XYZ on Twin Earth. But if Fred might be using "water" in a way that means the sentence is true in his mouth, what would make it the case that he isn't in fact using

the word "water" in this way? A plausible answer is that he might be using "water" for the watery kind he has mostly been in contact with. The Tuesday sentence would then get to be false by virtue of the watery kind he sees on that day is not (unknown to him) the kind he has mostly been acquainted with and is not the kind there is plenty of in the lake he is looking at on Tuesday. An implication of this answer is that if he stays on Twin Earth long enough, the reference of "water" in his mouth will shift to XYZ—as many hold. This is because there will come a time at which the watery kind he is mostly acquainted with is XYZ. But this account of why the shift happens is fully consistent with "water" being a word for telling about the watery stuff—the watery stuff that he's been mostly acquainted with up to the time he utters the sentence or uses the word.

The second reply that one might make to the argument is to argue that there are cases that tell us (6) is not a priori.[18] We can describe possible investigations of the nature of the actual world (our world) that we would rightly describe as showing that (6) is false. I think this is true—or rather that it might be true for the way some person or other uses the word "water"—but misses the key point. Let me spell this out.

Suppose someone describes a case in which it turns out that there are two watery kinds and argues that, in that case, we would think of water as akin to jade. We would not say there is no water. This means, he concludes, that (6) is not a priori; it is not a priori that water is *the* watery kind.

I reply, Where is the warrant for the "we"? Maybe some people use the word "water" in such a way that if it turned out that there were two watery kinds, they would say that there was no water. The objector should not assume that what he himself would say is what everyone ought to say. But more importantly, uniqueness is not important for the argument. Suppose someone uses the word "water" in a more inclusive way that does not require uniqueness; then we develop the key line of argument by replacing (6) by the following:

6*. Water is a watery stuff

for, on this usage, it is (6*) and not (6) that is a priori,
and we replace (7) by

7*. H_2O is the only watery stuff.

What is true in this case is that the inference from (4) and (7*) to (5) is a priori. We have, as before, an a priori passage from true sentences about how things are framed in terms of "H_2O" to a conclusion about how things are framed in terms of "water." The substitution of (7*) for (7) has no effect, for both are true.[19]

Or again, drawing on our discussion of Fred on Twin Earth, one might argue that we, or some of us, use "water" to talk about the watery stuff we are mostly acquainted with, saying that for us what's a priori is something like

6**. Water is the watery stuff we are mostly acquainted with.

In that case, (7) and (7*) would be replaced by

7**. H_2O is the watery kind we are mostly acquainted with

and what would then be true is that inference from (4) and (7**) to (5) is a priori. This time the point to note is that (7**) is true.

The crucial point is this. It would be wrongheaded to worry that the chemists who, in the late 1700s and early 1800s, carried out the experiments that showed that water is H_2O did one experiment too few. They forgot—amazingly—to do the experiment that showed that the watery stuff is water! The information they had when they did the experiments was enough to assure them that they were experimenting on water. If it hadn't been enough, we should be asking our chemistry departments to redo the experiments as a matter of urgency, worrying that the new experiments will give the "wrong" results and complaining about a major methodological lacuna that has escaped the notice of historians of science. But no one thinks we should be doing that, surely. However, the information they had back then did not include that the stuff they were experimenting on was H_2O. It was all to do with the stuff they were experimenting on being watery, on one or another way of spelling this out. Ergo, there is a way of spelling "watery"—maybe (plausibly?) in a disjunctive way that varies somewhat from one user of the word to another and that is suitably vague—that makes "Water =…watery…" a sentence that does not need experimental verification. Call it (6?). The crucial point is that, corresponding to (6?), there will be a (7?) that, when conjoined with (4), leads a priori to (5). Mutatis mutandis for other examples—that is, sentences other than (5).

7 A Doctrine about Concepts

Nowadays, a posteriori physicalism is often expressed in terms of concepts. The core idea common to many versions seems to be that having experiences with "feel" or a phenomenology—something's looking red to one, feeling a twinge in one's knee, and all that—enables one to have new sorts of thoughts.[20] How so? The idea seems to be that the very having of the experiences causes one to have new concepts—phenomenal concepts—and thereby new thoughts that deploy those concepts.[21] These thoughts are about the purely physical in the sense that phenomenal concepts apply to or refer to purely physical states, but they differ in a crucial respect from a thought like "the object before me is round." The concept of being round deployed in that thought applies to or refers to a purely physical property (being round); what is more, anyone who possesses that concept knows this (or maybe is in a position to know this without carrying out an experiment). This is not true for the phenomenal concepts. They do in fact refer to the purely physical, but this is not something one can know simply by possessing or grasping them. If one could, everyone would be a physicalist!

How might this help blunt the appeal of the knowledge and zombie arguments? Take Mary first. The idea is that there are thoughts she cannot have while in the black and white room. She does not have the concepts needed to have those thoughts. This means she has new thoughts when she leaves the room. What is more, because grasping the concepts deployed in those thoughts does not in itself reveal that their references are purely physical, she (or perhaps we should say, some of the commentators on the knowledge argument) make the mistake of thinking that the new thoughts are about new properties. Supporters of the zombie argument make a similar mistake. There are thoughts one can have only if one possesses and suitably deploys phenomenal concepts. But having these thoughts does not in itself reveal that the concepts deployed in them refer to the physical. This can tempt one into thinking that it is possible to duplicate with respect to the physical without duplicating with respect to what these thoughts refer to or are about. All the same, the thoughts *are* about the physical. What is more, the thoughts are *of course* about phenomenal conscious states; hence duplication with respect to the physical ensures duplication with respect to consciousness, but this fact is in a sense opaque. The opacity tricks supporters of the zombie argument into thinking that physical duplication need not ensure the presence of what the concepts refer to.

Here is what worries me about the story I have just sketched. What is the point of having concepts? The answer is that without them we could not think that things are thus and so for the relevant thus and so. Thinking that things are a certain way is bringing those things under a concept. Our thought is true if things are the appropriate way and false otherwise. Having the thought that something is a horse requires that I have the concept of a horse, and the thought will be true just if the something in question falls under the concept. Or consider lecturers in environmental science. They seek to impart the concept of a carbon sink to their students, and they do this in order that the students can have thoughts about which things are, and which things are not, carbon sinks. But having these thoughts would be of little value unless acquiring the concept of a carbon sink involves, as an essential component, knowing what it takes to be a carbon sink—knowing, that is, the reference of the concept. If the students don't know what it takes to be a carbon sink, they would be unable to have sensible opinions on how likely it was that something was a carbon sink or the implications of something's being a carbon sink.[22] The same goes for possessing the concept of a horse.

From this perspective on concepts,[23] the idea that we get a new concept, a so-called phenomenal concept, when we have certain experiences, a concept that is such that its possession is independent of knowing what it refers to, is an unattractive one. It goes against the whole point of possessing concepts. And it is worth emphasizing that the difficulty just raised doesn't rest on what some will think of as an unduly demanding notion of what it is to know the reference of a concept. Physicalists hold that *every* property of an experience is physical (they aren't closet attribute dualists). We

could allow that having the concept of a certain kind of experience only required a rather minimal knowledge of what it takes to be the experience in question, for the "rather minimal" knowledge will have to be knowledge of physical properties alone. For, according to physicalism, there aren't any other properties to have knowledge of.

I know, of course, that supporters of the special nature of phenomenal concepts give reasons designed to reconcile one to the idea that the reference of these concepts is especially opaque. I look at two of these reasons (both of which seem to me to be interesting in their own right).

Here's the first, put in the mouth of an imagined objector. "Egocentric concepts give the lie to the doctrine that grasping a concept involves knowing its reference. It is widely acknowledged that I can know all there is to know about how the world is and yet not know, as it might be, that I am the person who is spilling the sugar in the supermarket. I grasp the concept of being the person who is spilling the sugar—how else could I have, for example, the (false) thought, 'It's not me who is spilling the sugar'—but don't know its reference."[24]

I agree that I can know all there is to know about *how the world is* without knowing that I am the person spilling the sugar. I cannot, however, know all there is to know about *how I am* without knowing that I am the person spilling the sugar. Grasping egocentric concepts involves grasping how an individual has to be to fall under the concept. The example reinforces the message that grasping a concept involves grasping what it takes to fall under the concept. What the sugar and like examples tell us is that knowing how an individual is, is not the same as knowing how the world that individual is in is; they do not tell us that grasping a concept floats free of knowing its reference.

The way this point plays into debates over the nature of the inference from the physical to the mental is that it reminds us that there are two topics to distinguish. One is the nature of the inference from the physical way the *world* is to the mental way the *world* is. The other is the inference from the physical way an *individual* is to the mental way that *individual* is (the way an individual is does not of course mean the intrinsic way an individual is—much of how I am is a matter of how I am related to the world).

Here's the second, put in the mouth of an imagined objector. "Phenomenal concepts have a special demonstrative, recognitional nature. They are 'It's that again' concepts. When something looks red to me here and now, I can, so to speak, internally demonstrate the state I am in and fix it in my mind. I am then able to recognize it when I am again in it and can distinguish it from the state I am in when something looks green to me. Phenomenal concepts are, in this sense, acquaintance concepts and not descriptive concepts, and this is why grasping a phenomenal concept does not deliver how something has to be to fall under it. That thesis is only true for descriptive concepts."

I agree that there is a demonstrative-recognitional element in having an experience. As Wittgenstein (1963, sec. 610) said, it is very hard to describe the aroma of coffee: "Describe the aroma of coffee.—Why can't it be done? Do we lack the words?" We can,

however, recognize the aroma and can point to it in the sense of making demonstrative reference to it. Indeed, what else is going on when your sense of smell is being tested and you are required, as it might be, to press a button when you smell coffee? However, the claim about description is a point about words. It is not a point about kinds. Any suggestion that smelling coffee is the same kind of experience as something's looking red to me should surely be dismissed out of hand. They are very different *kinds* of experiences, and when we exercise our recognitional capacity with respect to them, we recognize the difference and are aware of the kind. The same goes for when we make demonstrative reference to our current experiences. We point to different kinds of experiences and know that we are pointing to different kinds of experiences. When, in a psychology experiment, we agree to press a button once when we smell coffee (as coffee) and twice when we see something red (as red), we know what we are agreeing to do and are confident that we will do a reasonable job of complying.

A fan of diaphanousness or transparency may insist (as I would) that the nature of the experience is the putative nature of what's experienced and that the role sometimes given to introspection in revealing the kind in question is at best misleading. One looks out, as we might put it, when latching onto the aroma of coffee or the redness of a sunset: the kind is the kind the thing one is aware of putatively belongs to, and the demonstrative reference—the pointing to—we are talking about is to the kind that we are perceiving putatively belongs to.[25] None of this alters the fact that we are dealing with kinds: grasping the concepts involves inter alia grasping the kinds in question. The fan of diaphanousness should not be read as denying that; rather, that person has an account of the kinds in question—they are the kinds the thing we are aware of putatively belong to.

Here is a simple example to highlight the point.[26] Some people are motion blind. They have the concept of motion: they know what it takes for something to move—it needs to occupy different places at different times—they know, that is, the concept's reference. However, they cannot see something *as* moving. There is a certain kind of experience we can enjoy and that they cannot. As supporters of phenomenal concepts might say it, there is a phenomenal concept that we have and they lack. If we sought to tell someone suffering from akinetopsia what it is like to see something as moving, we would find ourselves in the aroma-of-coffee situation. We can recognize and demonstrate to ourselves the experience in question, but the words we produce do not convey to the motion blind the nature of the experience that we sometimes have and they never have. However, it would be wrong to say that when we have the distinctive experience of seeing something as moving that we are deploying a concept whose application conditions are opaque to us. We know what it takes for something to be as we experience it to be when we see it as moving: it needs to be in different places at different times.[27] No amount of talk about recognition, demonstration, experiential nature, what it is like to see something as moving, the role of acquaintance, the impossibility

of conveying to the motion blind what it is like to see something as moving, and so on, should blind us to this.

These remarks leave us with an obvious question. There is a big difference between the motion case and, for example, the color case. The second case but not the first invites knowledge argument type thoughts. Seeing something as colored does seem to involve awareness of a new sort of property; seeing something as moving does not. This is why I said earlier that the "what it's like" rubric is potentially misleading as a way of capturing the phenomenal aspect of seeing something as colored. In both the color case and the motion case there is something it is like: in the one to see something as moving, in the other to see something as red, but what they share isn't the source of the challenge posed by mental states with a phenomenology. It is what *divides* them that provides the challenge.

What's the explanation of the big difference between the two cases? That question occupies us in the short final section. We will see that the answer throws light on the intuitive appeal of a posteriori physicalism.

8 The Qualia Illusion

Pains in phantom limbs are awfully real. If you have a severe one, it is awfully tempting to describe the situation in terms of your standing in a kind of immediate acquaintance relation to a pain located in, say, the space where your foot would once have been. It is now widely agreed that this picture is seriously flawed.

Here is a second example. Sometimes we seem to be conscious of exactly when we make a decision. Perhaps we tell an experimenter just where a rotating pointer is when we make the decision to lift a finger. It turns out that the time we report is too late. The brain sets things in motion well before then (by neuroscience standards). But this is not an argument for epiphenomenalism about conscious decisions, as is widely agreed. What we learn from the experimental results, in my unoriginal view, is that what presents as the time of the conscious decision is in fact the time we represent the decision as taking place.[28]

In both these cases it is natural to talk of illusion. Things are not quite as they seem. This is something we have learned to say about the kinds of visual phenomena that fueled the famous or infamous argument from illusion. The idea that we stand in some kind of acquaintance relation to two curved lines when presented with the Hering illusion—two curved lines in one's visual field, as it is natural to say it—is a very attractive one. There was always a lot going for the sense-datum theory. But most of us now agree that this cannot be the right way to think of what's going on.[29] We quarrel about what to replace the sense-datum theory with but agree it needs replacement.

I think color vision is another case in which appearances mislead. When something looks red to us, it seems that we are in a kind of acquaintance relation to an instance

of a highly distinctive property, a property that in some ways is like the aroma of coffee. Words aren't much good for describing it to someone who cannot see something as red. The difference between seeing something as red and seeing something as moving isn't that, in the second case but not the first, we know the property we are seeing something as having. In both cases, we see the object as being a certain way: the experience of color is not the experience of "blank."[30] Indeed, it is hard to think of anything less blank-like than seeing something as red. The difference is that the property we see something as having when it looks red to us is a property it does not have, and in fact nothing in our world has the property.

Why do I say that nothing has the property that something looks to have when it looks red? The properties that objects look to have when they look red, green, blue, yellow, and so on, are intrinsic properties of the objects that *essentially* stand in various similarity and difference relations. The color we see something to have when it looks red to us is essentially similar to the color something looks to us to have when it looks pink. It is a feature of the properties themselves and not, for instance, of the way the properties impact on creatures with eyes and brains like ours. I think it is a discovery about the nature of the world we live in that no such properties are anywhere instantiated in it.

I think the intuition that drives a posteriori physicalism is the correct thought that we could never deduce the presence of one of these properties from the physical way things are. (We had better not be able to make the deduction, add physicalists.) In discussion, people insist that we have to allow that what it's like to see something as red, green, and so on, is instantiated.[31] Isn't that close to common ground, and doesn't that imply that the properties in question are instantiated? But what it is like to see something as moving can be instantiated when nothing is moving. Moreover, what I am talking about here is, as I say above, a point of *difference* between seeing something as moving and seeing it as red. That there is something it is like to see something as moving and something it is like to see it as red is a point in *common* between the two cases, as is the point that, in both cases, we have a hard time telling the motion blind and the color blind what it is like.[32]

How serious is the illusion associated with color experience? Should we say that it means nothing is colored? Strictly speaking, yes, if what I say above is right. Nothing is as the experience of looking colored represents it to be. But I grant, in common with many attracted by eliminativism about color, that it is sensible to set lower standards in everyday chat. Although nothing in our world is exactly as our experience of color represents things to be, the extent of the divergence will vary case by case, and when the divergence is at a minimum by the standards set by the way things actually are, it is sensible to speak, as it might be, of something as being red.[33]

Notes

1. See, e.g., Smart (1959) and Armstrong (1968). Much of Armstrong's book is devoted to offering analyses of mental states, in a more or less traditional sense of "analysis." Interestingly, in the preface to the revised edition (1993, xv), he expresses reservations on just this point.

2. The *A*-type versus *B*-type terminology is due to David Chalmers (2012, 341). I prefer the a priori versus a posteriori terminology; it seems to me to be more transparent.

3. The chemistry of water is complex, as many have noted. It would have been better to use, say, gold and kind with atomic number 79, saying, as it might be, that the way the distribution of stuff with atomic number 79 necessitates without a priori entailing the distribution of gold is a model for understanding the relation between the physical way things are and the mental way things are. I, however, bow to precedent and use the water-H_2O example.

4. See, e.g., Jackson (1994) and Block and Stalnaker (1999). (Despite my recalcitrance, as Block will see things, I am very much indebted to him.)

5. Or if you think the debate is really a semantic one about how best to use the term "information," we can specify the best way to use the term without using the term itself (fortunately, as otherwise we'd have a vicious circularity).

6. But note that this is a much more inclusive use of "physical" than that which typically figures in supervenience explications of physicalism.

7. After all, a priori and a posteriori physicalists alike have to grant that there's a viable notion of a physical property in the offing.

8. Thanks here to Eden Lin.

9. They may well go on to insist that there are examples (especially in mathematics) where there *is* a change in truth value on substituting sentences that are a priori equivalent in belief contexts.

10. But see, e.g., Jackson (2010a, Lecture four).

11. The mistake here would be like that some accuse Quine (1961) of making in his attack on essentialism. For some thoughts on whether Quine did make this mistake, see Cartwright (1968) and Jackson (2010b).

12. Near the end of this chapter, I argue that the intuitions that drive the zombie and knowledge arguments concern *uninstantiated* properties, in which what is uninstantiated are not the properties I am talking about here but rather those that the experiences represent things as having.

13. The presentations in Jackson (1982, 1986), however, are in terms of physical information, in which that isn't defined by the words that deliver the information.

14. I am using "entail" in place of "determine" when sentences and words are the focus. Some use "entail" to mean what I use "a priori entail" for. Usually they are thinking of entailment as the converse of a priori derivability.

15. Others may prefer to phrase the preceding paragraph in terms of reference fixing in Kripke's sense, the key claim then being that "the watery stuff" (or being the watery stuff) is the reference fixer for "water."

16. This seems to be Stalnaker's position (2003, esp. 196).

17. Of a kind with the version discussed in Stalnaker (2008, 116) in his discussion of Boghossian (1994).

18. My sense is that this is the most popular one. For one example, see Block and Stalnaker (1999).

19. Mutatis mutandis if it is indeterminate or changes over time which of (7) or (7*) is true.

20. The "seems" in this sentence signals that I am offering one way of construing a set of views, sometimes badged "the phenomenal concepts strategy"; see, e.g., Loar (1990) and the discussion in Block (1994, 515). I am indebted to discussions with many, including especially Helen Yetter Chappell.

21. This is the typical way of acquiring the new concepts, but having a neurosurgeon make certain changes to one's brain (those associated with seeming to remember a twinge in one's knee, e.g.) would be another, as nearly all agree.

22. Well, not quite. If ignoramus hears someone he trusts say, "The ocean is a carbon sink," he may know that the ocean is a carbon sink in the sense that he knows the sentence in English is true. But he doesn't know how the ocean is being said to be. The division of linguistic labor (Putnam 1975) doesn't mean that ignoramus passes the environmental science exam.

23. Philosophers talk about concepts a lot. I do not pretend to be capturing what everyone has in mind.

24. The example comes from Perry (1979).

25. Of course, I am capable of reflecting on how I putatively stand with respect to the world around me or how I stand with respect to a part of my body, at a given moment, and that involves looking inward in some good sense. I attend to a relation between myself and the way things putatively are. Fans of diaphanousness do not, or need not, deny the existence of introspection.

26. Suggested to me by Pettit (2009).

27. The metaphysics of something's being in different places at different times is a matter of debate between, e.g., four-dimensionalists and three-dimensionalists about time, but we can reasonably set that debate aside here.

28. See, e.g., Dennett's (1991, 162–163) discussion of Benjamin Libet's experiments.

29. Including a one-time supporter of the sense-datum theory like me.

30. *Pace* Armstrong (1968, chap. 12) but see the clarification in the final endnote.

31. Much of it is in the course of objecting to Jackson (2003).

32. This is where, in my view, the ability reply to the knowledge argument—Nemirow (1990) and Lewis (1990)—falls short. It seems right that what the motion blind lack is an ability and that learning what it is like to see something as moving will be the acquisition of an ability. But the property we see something as having when we see it as moving is a property we know is sometimes instantiated and is a property that transparently follows a priori from the physical way the world is. If what I say above is right, this is precisely not the case for seeing something as red, green, and so on.

33. Nowadays I am not nearly as strict. There plausibly are properties instantiated in our world that stand in the various similarity and difference relations I talk of in the text. The illusion is rather that these properties are revealed to us when something looks a certain color. Of course, what's important for the argument in the text is that there is an illusion, for that's what does the work in explaining the intuition that drives a posteriori physicalism.

References

Armstrong, D. M. 1968. *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul. Revised edition 1993.

Block, N. 1994. Qualia. In *A Companion to the Philosophy of Mind*, S. Guttenplan, 514–520. Oxford: Blackwell.

Block, N., and R. Stalnaker. 1999. Conceptual analysis, dualism and the explanatory gap. *Philosophical Review* 90: 5–43.

Boghossian, P. 1994. The transparency of mental content. *Philosophical Perspectives* 8: 5–26.

Cartwright, R. 1968. Some remarks on essentialism. *Journal of Philosophy* 65: 615–626.

Chalmers, D. J. 2012. *Constructing the World*. Oxford: Oxford University Press.

Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.

Jackson, F. 1982. Epiphenomenal qualia. *Philosophical Quarterly* 32: 127–136.

Jackson, F. 1986. What Mary didn't know. *Journal of Philosophy* 83: 291–295.

Jackson, F. 1994. Finding the mind in the natural world. In *Philosophy and the Cognitive Sciences, Proceedings of the 16th International Wittgenstein Symposium*, R. Casati, B. Smith, and G. White, 101–112. Vienna: Hölder-Pichler-Tempsky.

Jackson, F. 2003. Mind and illusion. In *Minds and Persons: Royal Institute of Philosophy Supplement 53*, A. O'Hear, 251–271. Cambridge: Cambridge University Press.

Jackson, F. 2010a. *Language, Names, and Information*. Oxford: Wiley-Blackwell.

Jackson, F. 2010b. Possible worlds and the necessary *A Posteriori*. In *Modal Content and Modal Knowledge: Essays on the Metaphysics and Epistemology of Modality*, B. Hale and A. Hoffman, 257–266. Oxford: Oxford University Press.

Kripke, S. 1980. *Naming and Necessity*. Oxford: Basil Blackwell.

Lewis, D. 1990. What experience teaches. In *Mind and Cognition*, W. Lycan, 499–519. Oxford: Basil Blackwell.

Loar, B. 1990. Phenomenal states. *Philosophical Perspectives* 4: 81–108.

Nemirow, L. 1990. Physicalism and the cognitive role of acquaintance. In *Mind and Cognition*, W. Lycan, 490–499. Oxford: Basil Blackwell.

Papineau, D. 2011. Phenomenal concepts and the private language argument. *American Philosophical Quarterly* 48: 175–184.

Perry, J. 1979. The problem of the essential indexical. *Noûs* 13: 3–21.

Pettit, P. 2009. Consciousness and the frustrations of physicalism. In *Minds, Ethics, and Conditionals*, Ian Ravenscroft, 163–188. Oxford: Clarendon Press.

Putnam, H. 1975. The meaning of "meaning." In *Mind, Language and Reality*, 215–271. Cambridge: Cambridge University Press.

Quine, W. V. 1961. Reference and modality. In *From a Logical Point of View*, 139–159. New York: Harper.

Smart, J. J. C. 1959. Sensations and brain processes. *Philosophical Review* 68: 141–156.

Stalnaker, R. 2003. On considering a possible world as actual. In *Ways a World Might Be*, 188–200. Oxford: Clarendon Press.

Stalnaker, R. 2008. *Our Knowledge of the Internal World*. Oxford: Clarendon Press.

Wittgenstein, L. 1963. *Philosophical Investigations*. Oxford: Basil Blackwell.

-1—
0—
+1—