

Oxford Handbooks Online

The Two-Dimensional Argument Against Materialism

David J. Chalmers

The Oxford Handbook of Philosophy of Mind

Edited by Ansgar Beckermann, Brian P. McLaughlin, and Sven Walter

Print Publication Date: Jan 2009 Subject: Philosophy, Philosophy of Mind

Online Publication Date: Sep 2009 DOI: 10.1093/oxfordhb/9780199262618.003.0019

Abstract and Keywords

A number of popular arguments for dualism start from a premise about an epistemic gap between physical truths and truths about consciousness, and infer an ontological gap between physical processes and consciousness. Arguments of this sort include the conceivability argument, the knowledge argument, the explanatory gap argument, and the property dualism argument. Such arguments are often resisted on the grounds that epistemic premises do not entail ontological conclusions. This article views that one can legitimately infer ontological conclusions from epistemic premises, if one is careful about how one reasons. To do so, the best way is to reason first from epistemic premises to modal conclusions (about necessity and possibility), and from there to ontological conclusions. Here the crucial issue is the link between the epistemic and modal domains.

Keywords: dualism, epistemic gap, physical truths, truths about consciousness, ontological gap, physical processes, consciousness

A number of popular arguments for dualism start from a premise about an epistemic gap between physical truths and truths about consciousness, and infer an ontological gap between physical processes and consciousness. Arguments of this sort include the conceivability argument, the knowledge argument, the explanatory gap argument, and the property dualism argument. Such arguments are often resisted on the grounds that epistemic premises do not entail ontological conclusions.

My view is that one can legitimately infer ontological conclusions from epistemic premises, if one is careful about how one reasons. To do so, the best way is to reason first from epistemic premises to modal conclusions (about necessity and possibility), and from there to ontological conclusions. Here the crucial issue is the link between the epistemic and modal domains. How can one reason from theses about what is knowable or conceivable to theses about what is necessary or possible?

The Two-Dimensional Argument Against Materialism

To bridge the epistemic and modal domains, the framework of two-dimensional semantics can play a central role. I have used this framework in earlier work (Chalmers 1996) to mount an argument against materialism. Here I want to revisit the argument, (p. 314) laying it out in a more explicit and careful form, and responding to a number of objections.

18.1 The Conceivability Argument

The most straightforward form of the conceivability argument against materialism runs as follows.

- (1) $P \& \sim Q$ is conceivable.
 - (2) If $P \& \sim Q$ is conceivable, $P \& \sim Q$ is metaphysically possible.
 - (3) If $P \& \sim Q$ is metaphysically possible, materialism is false.
-
- (4) Materialism is false.

Here P is the conjunction of all microphysical truths about the universe, specifying the fundamental features of every fundamental microphysical entity in the language of microphysics. Q is an arbitrary phenomenal truth: perhaps the truth that someone is phenomenally conscious, or perhaps the truth that a certain individual (that is, an individual satisfying a certain description) instantiates a certain phenomenal property. $P \& \sim Q$ conjoins the former with the denial of the latter.

If Q is the truth that someone is phenomenally conscious, then $P \& \sim Q$ is the statement that everything is microphysically as in our world, but no one is phenomenally conscious. In this version $P \& \sim Q$ says that the world is a *zombie world*. If Q is the truth that a certain individual instantiates a certain phenomenal property, then $P \& \sim Q$ is the statement that everything is microphysically as in our world, but that it is not the case that the individual in question instantiates the relevant phenomenal property. In this case it will suffice for the truth of $P \& \sim Q$ that the world is a *zombie world*, or simply that the individual in question is a *zombie* in a physically identical world. It will also suffice that the individual in question is an *invert*, who has an experience that differs slightly from the corresponding experience of the corresponding individual in our (physically identical) world.

The first premise of this argument asserts an epistemic thesis, about what can be conceived. The second premise steps from the epistemic thesis to a modal thesis, about what is possible. The third premise steps from the modal thesis to a metaphysical thesis, about the nature of our world.

The third premise is relatively uncontroversial. It is widely accepted that materialism has modal commitments. Some philosophers question whether materialism is equivalent to a modal thesis, but almost all accept that materialism at least *entails* a modal thesis. Here

The Two-Dimensional Argument Against Materialism

one can invoke Kripke's metaphor: if it is possible that there is a world physically identical to our world but phenomenally different, then after God fixed the physical facts about our world, he had to do more work to fix the phenomenal facts.

A familiar complication arises from the observation that physicalism about our world is compatible with the possibility of dualism in other worlds, and in particular (p. 315) is compatible with the possibility of a physically identical world that contains extra, non-physical phenomenology. This means that if Q is a negative truth about our world—say, the truth that no one instantiates a certain phenomenal property—then materialism about our world is compatible with the possibility of $P \& \sim Q$. To finesse this point we can stipulate that in the argument above, Q is a positive truth (one that holds in all worlds that contain a duplicate of our world: see Chalmers 1996: 40): if Q is a positive truth, then materialism is incompatible with the possibility of $P \& \sim Q$. Alternatively, we can conjoin P with a 'that's-all' statement T , stating that the world is a *minimal* world that satisfies P (see Jackson 1998: 26). Then even when Q is a negative truth, materialism is not compatible with the possibility of $PT \& \sim Q$ (where PT is the conjunction of P and T).

The real work in the argument is done by the first and second premises. The second premise is particularly controversial, as there are a number of examples that have led many philosophers to deny that there is an entailment from conceivability to metaphysical possibility. To assess these premises we need to understand the notion of conceivability.

18.2 Varieties of Conceivability

Conceivability is to be understood as an epistemic notion, defined in epistemological (and perhaps psychological) terms. To a first approximation, we can say that S is conceivable when S expresses a coherent hypothesis: one that cannot be ruled out a priori. To refine this understanding it is useful to make some distinctions. (These distinctions are discussed at much greater length in Chalmers 2002.)

We can say that S is *prima facie* conceivable for a subject when that subject is unable to rule out the hypothesis expressed by S by a priori reasoning, on initial consideration. We can say that S is *ideally* conceivable when the hypothesis expressed by S cannot be ruled out a priori, even on ideal rational reflection. The main difference here is that *prima facie* conceivability is tied to a subject's contingent cognitive limitations, while ideal conceivability abstracts away from those limitations.

Some examples: (1) ' $2 + 2 = 5$ ' is neither *prima facie* conceivable nor ideally conceivable. (2) Where S is a highly complex but provable mathematical truth, $\sim S$ will be *prima facie* conceivable for most subjects, but it is not ideally conceivable. (3) Where S is 'There is a flying pig', S is *prima facie* conceivable, and is almost certainly ideally conceivable.

The Two-Dimensional Argument Against Materialism

The notions of conceivability discussed above are versions of *negative* conceivability, which is defined in terms of what a subject can *rule out* through a priori reasoning. We can say that S is negatively conceivable when S cannot be ruled out through a priori reasoning. The two notions above can then be seen as *prima facie* negative conceivability and ideal negative conceivability respectively.

It is also possible to define notions of *positive* conceivability, which is defined in terms of what subjects can form a positive conception of. We can say that S is (p. 316) positively conceivable when one can coherently imagine a situation in which S is the case. Where negative conceivability requires merely entertaining a hypothesis and being unable to rule it out, positive conceivability involves being able to form some sort of clear and distinct conception of a situation in which the hypothesis is true. One can then say that S is *prima facie* positively conceivable when a subject can imagine a situation that they take to be coherent and that they take to be one in which S is the case. And one can say that S is *ideally* positively conceivable when its *prima facie* positive conceivability cannot be defeated on ideal rational reflection (in particular, when arbitrary details can be filled in in the imagined situation without any contradiction revealing itself, and when ideal reflection reveals the imagined situation as one in which S is the case).

Traditional notions of conceivability (Descartes's clear and distinct conceivability, for example) are arguably varieties of positive conceivability rather than negative conceivability. At the same time, the notion of positive conceivability is more complex than that of negative conceivability, and a rigorous characterization of the notion requires saying much more about just what it is to imagine a situation, and so on. I characterize positive conceivability in more depth in Chalmers (2002). In this chapter the roles of positive and negative conceivability will often be interchangeable (I will make clear when the difference is relevant), so the informal account above will suffice for present purposes. For much of the discussion one can focus on negative conceivability without much loss, but positive conceivability is available as an alternative if there turn out to be any problems with theses tied to negative conceivability.

In so far as there is a gap between *prima facie* conceivability and ideal conceivability, it is ideal conceivability that is a better guide to possibility. This is especially clear in the case of *prima facie* negative conceivability: we have seen that the negation of a complex mathematical truth may be *prima facie* negatively conceivable, but it is not ideally conceivable and it is not possible. It is less easy to find cases of *prima facie* positive conceivability without ideal positive conceivability (see Chalmers 2002 for some potential cases), but in so far as there are such cases, there will be little reason to think that they are possible.

Correspondingly, some familiar purported counter-examples to the claim that conceivability entails possibility are really counter-examples to the claim that *prima facie* conceivability entails possibility. For example, it is sometimes said that both Goldbach's conjecture and its negation are conceivable, while only one of them is possible. Here the relevant notion of conceivability is something like *prima facie* negative conceivability.

The Two-Dimensional Argument Against Materialism

There is no reason to believe that both Goldbach's conjecture and its negation are ideally conceivable, so there is no reason to think that this is a counter-example to the claim that ideal conceivability entails possibility. So from here onward talk of 'conceivability' *simpliciter* should always be understood to be talk of ideal conceivability (either positive or negative).

The other familiar class of purported counter-examples arises from Kripke's analysis of the necessary a posteriori. It is often said that sentences such as 'Water is not H₂O' provide counter-examples to the claim that conceivability entails possibility: it is conceivable that water is not H₂O, but it is not metaphysically possible.

(p. 317)

Here one has to be careful. There is a *sense* of 'conceivable' in which 'Water is not H₂O' is not conceivable (given that water is H₂O in the actual world): in this sense, any conceivable situation in which it seems that water is not H₂O (a Twin Earth world, say) would better be described as a conceivable situation in which water is still H₂O, but in which there is watery stuff that is not H₂O. Using the term 'conceivable' this way, one might say that 'Water is not H₂O' seems conceivable (or that it is *prima facie* conceivable, to one without relevant empirical knowledge), but that it is not really conceivable. We might call this sense of conceivability *secondary conceivability* (for reasons familiar from a two-dimensional analysis and discussed in Chalmers 2002). Then the Kripkean cases are compatible with the claim that secondary conceivability entails metaphysical possibility. But at the same time this claim is not very useful for present purposes, as whether a sentence is secondarily conceivable will typically depend on a variety of empirical factors, and an opponent might deny that zombies are secondarily conceivable, on the grounds that there is an a posteriori identity between phenomenal and physical properties. So a link between secondary conceivability and possibility does not offer an a priori route to conclusions about metaphysical possibility.

Instead, what is relevant here is *primary conceivability*: the sense in which 'Water is not H₂O' can correctly be said to be conceivable. The notion of negative conceivability defined above is a sort of primary conceivability, as it is defined in terms of what can be ruled out a priori, and 'Water is H₂O' cannot be established a priori. (One might define a distinct notion of negative secondary conceivability, but I will set that aside here.) One can likewise define a notion of positive primary conceivability, so that S is positively primarily conceivable when a subject can imagine a coherent situation that verifies S, where a situation verifies S when, under the hypothesis that the situation actually obtains, the subject should conclude that S. If the subject imagines a Twin Earth situation with XYZ in the oceans and lakes, and assumes that the situation obtains in their own environment, then the subject should conclude that water is XYZ rather than H₂O. So 'Water is not H₂O' is positively primarily conceivable, as well as negatively primarily conceivable.

The Two-Dimensional Argument Against Materialism

Unlike secondary conceivability, matters of primary conceivability are plausibly in the a priori domain: whether S is primarily conceivable turns on matters of a priori reasoning. But primary conceivability does not entail metaphysical possibility: 'Water is not H₂O' is primarily conceivable, but it is not metaphysically possible.

Still, there remains a link between primary conceivability and metaphysical possibility in these cases. When we conceive that water is not H₂O we imagine (for example) a Twin Earth situation in which the watery liquid in the oceans and lakes is XYZ. This situation is metaphysically possible, so there is a sense in which our conceiving involves access to a possible world. Under the usual way of describing possible worlds, this world is not a world in which water is not H₂O. But the world still stands in a strong relation to the sentence 'Water is not H₂O.' In particular, if we came to accept that our own world had the character of this world (with XYZ in the oceans and lakes), we should then endorse the claim 'Water is not H₂O.'

(p. 318)

This can be put in two-dimensional terms by saying that while the Twin Earth does not *satisfy* 'Water is not H₂O' ('Water is not H₂O' is not true of that world considered as counterfactual), the Twin Earth world *verifies* 'Water is not H₂O' ('Water is not H₂O' is true of that world considered as actual). Equivalently, we can say that while the *secondary intension* of 'Water is not H₂O' is false at W, the sentence's *primary intension* is true there. To a first approximation, a world W verifies S (or S is true at W considered as actual, or the primary intension of S is true at W) when if we came to accept that our own world is qualitatively like W we should then endorse S. Strictly speaking, the worlds W that are relevant to primary intensions are *centred* worlds: worlds that come with a marked 'centre' consisting of an individual and time. When we consider a centred world W as actual, we consider the hypothesis that we are currently in the situation of the individual at the centre. (For much more on these notions see Chalmers 2004.)

We can say that when the primary intension of S is true at some centred world (i.e. when some centred world verifies S) S is *primarily possible*, or *1-possible*. When the secondary intension of S is true at some world (i.e. when some world satisfies S), S is *secondarily possible*, or *2-possible*. Then 'Water is not H₂O' is not 2-possible, but it is 1-possible.

The observation that sentences such as 'Water is not H₂O' are conceivable but not possible, in these terms, comes to the claim that these sentences are primarily conceivable (1-conceivable) but are not secondarily possible (2-possible). So there is good reason to believe that 1-conceivability does not entail 2-possibility. However, these cases are entirely compatible with a link between 2-conceivability and 2-possibility, and, more importantly for present purposes, they are entirely compatible with a link between 1-conceivability and 1-possibility.

In fact, it is not hard to argue that all of the standard Kripkean a posteriori necessities ('Heat is the motion of molecules', 'Hesperus is Phosphorus', and so on) have this structure. For each of these necessities one might say that its negation is conceivable but

The Two-Dimensional Argument Against Materialism

not possible, meaning that it is 1-conceivable but not 2-possible. But in each of these cases the sentence in question is 1-possible. For example, 'Heat is not the motion of molecules' is verified by a world in which something other than molecules causes sensations as of heat. 'Hesperus is not Phosphorus' is verified by a world in which the objects visible in the morning and evening skies are entirely distinct. Furthermore, it is plausible that worlds such as these are just what one is conceiving of when one conceives that heat is not the motion of molecules, or that Hesperus is not Phosphorus. So in these cases there remains a strong link between conceivability and metaphysical possibility.

To summarize: we have seen that the standard counter-examples to a conceivability-possibility link are accommodated by noting that (i) *prima facie* conceivability is an imperfect guide to possibility, and (ii) primary conceivability is an imperfect guide to secondary possibility. But (i) is entirely consistent with a link between ideal conceivability and possibility, and (ii) is entirely consistent with a link between primary conceivability and primary possibility. Putting the pieces (p. 319) together: all of these counter-examples are compatible with the thesis that ideal primary conceivability entails primary possibility.

There are two versions of this thesis, depending on whether one interprets the relevant sort of conceivability as positive or negative.

(CP+) Ideal primary positive conceivability entails primary possibility.

(CP-) Ideal primary negative conceivability entails primary possibility.

CP- entails CP+, as ideal primary positive conceivability entails ideal primary negative conceivability. If S can be ruled out a priori, then no coherent imagined situation will verify S. It is not obvious whether or not CP+ entails CP-, as it is not obvious whether ideal primary negative conceivability entails ideal primary positive conceivability. That is, it is not obvious whether or not there is an S that cannot be ruled out a priori, but such that no coherent imagined situation verifies S. (In Chalmers 2002 I argue that there is no such S, so that ideal primary negative conceivability entails ideal primary positive conceivability.) So CP- is at least as strong as CP+ and is possibly somewhat stronger.

Most importantly for present purposes, however, both CP+ and CP- are compatible with all the familiar purported counter-examples to the conceivability-possibility link.

Furthermore, it seems that there are no clear counter-examples to either thesis (though later in the chapter I will discuss some potential counter-examples that have been put forward). In particular, both theses are entirely compatible with the existence of Kripkean a posteriori necessities, so while existence of these necessities is often used to cast doubt on conceivability-possibility theses, they cannot be used to cast doubt on CP+ or CP-.

So for now I will take these theses as reasonable conceivability-possibility theses that might be used in mounting a refined conceivability argument against materialism. Later in the chapter I will return to the question of their truth.

18.3 A Refined Conceivability Argument

Henceforth, unqualified uses of 'conceivability' and 'conceivable' should be understood as invoking ideal primary conceivability. I will often be inexplicit about whether positive or negative conceivability is involved. In effect, the argument forms below can be understood as generating two different arguments, depending on whether one understands conceivability as ideal primary positive conceivability or as ideal primary negative conceivability. For many purposes the distinction will not matter. When it does matter, I will be explicit. Given the discussion above, one might try generating an anti-materialist argument by simply substituting primary possibility for metaphysical possibility in the original argument.

- (1) $P \& \sim Q$ is conceivable.
- (2) If $P \& \sim Q$ is conceivable, $P \& \sim Q$ is 1-possible.

(p. 320)

- (3) If $P \& \sim Q$ is 1-possible, materialism is false.

-
- (4) Materialism is false.

On this reading (1) and (2) are both plausible theses, but (3) is not obviously plausible. The reason is that materialism requires not the 1-impossibility of $P \& \sim Q$ but the 2-impossibility of $P \& \sim Q$. That is, materialism requires that it *could not have been the case* that P were true without Q being true. This is a subjunctive claim about ordinary metaphysical possibility, and so invokes 2-impossibility rather than 1-impossibility.

A materialist might reasonably question (3) by holding that even if there is a world W *verifying* $P \& \sim Q$, W might be a world with quite different ingredients from our own. For example, it might be that W does not instantiate true microphysical properties (those instantiated in our world), such as mass and charge, but instead instantiates quite different properties; say, pseudo-mass and pseudo-charge, which stand to mass and charge roughly as XYZ stands to H₂O. Likewise, it might be that W does not lack true phenomenal properties, but instead lacks quite different properties: say, pseudo-phenomenal properties. If so, then the possibility of W has no bearing on whether true microphysical properties necessitate true phenomenal properties. And it is the latter that is relevant for materialism.

Still, it may be that the gap between 1-possibility and 2-possibility could be closed. In particular, when a statement S has the same primary intension and secondary intension, then a world will verify S iff it satisfies S, so S will be 1-possible iff it is 2-possible. If P and Q both have primary intensions that coincide with their secondary intensions, then so will $P \& \sim Q$, and we could run the following argument:

- (1) $P \& \sim Q$ is conceivable.
- (2) If $P \& \sim Q$ is conceivable, $P \& \sim Q$ is 1-possible.

The Two-Dimensional Argument Against Materialism

(3) If $P \& \sim Q$ is 1-possible, $P \& \sim Q$ is 2-possible.

(4) If $P \& \sim Q$ is 2-possible, materialism is false.

(5) Materialism is false.

Here the truth of (3) requires that both P and Q have primary and secondary intensions that coincide. In the case of Q, this claim is quite plausible. As Kripke noted, there does not seem to be the same strong dissociation between appearance and reality in the case of consciousness as in the cases of water and heat: while it is not the case that anything that looks like water is water, or that anything that feels like heat is heat, it is plausibly the case that anything that feels like consciousness is consciousness. So it is not clear that the notion of 'pseudo-consciousness', something that satisfies the primary intension of 'consciousness' without being consciousness, is coherent. Likewise for other more specific phenomenal properties. So there is a strong case that the primary and secondary intensions of phenomenal terms coincide. (For more on this case see Chalmers 2003.)

However, in the case of P this claim is less plausible. A materialist might reasonably hold that microphysical terms (such as 'mass' and 'charge') have primary intensions that differ from their secondary intensions. In particular, it is plausible that the primary intensions of these terms are tied to a certain theoretical role. We might (p. 321) say that the primary intension of 'mass' picks out whatever property plays the mass role (e.g. resisting acceleration in certain ways, being subject to mutual attraction in a certain way, and so on), and that the primary intension of 'charge' picks out whatever property plays the charge role (e.g. obeying certain electromagnetic principles, being subject to attraction and repulsion in certain ways, and so on).

By contrast, one might reasonably hold that the secondary intension of microphysical terms is tied to the property that actually plays the role. For example, if property M plays the mass role in the actual world, then one might hold that in any world in which mass is instantiated, mass is M. It follows that if there are worlds in which some *other* property M' plays the mass role, then M' is not mass in that world (at best, it is pseudo-mass). If so, then the primary and secondary intensions of 'mass' will not coincide: the primary intension of 'mass' will pick out whatever plays the mass role in such a world, but the secondary intension will not.

There are other views of the semantics and metaphysics of microphysical terms that may reject this argument for the distinctness of the primary and secondary intensions of 'mass'. In particular, the argument will not go through on views according to which it is necessary that mass is the property that plays the mass role. (These include views on which 'mass' is a non-rigid designator whose secondary intension picks out different properties that play the role in some worlds, and views on which it is necessary that M is the property that plays the mass role, where M is the property rigidly designated by 'mass'.) Still, the view sketched above is a quite reasonable view—more plausible than the alternatives, in my opinion—and it is the view that best grounds resistance to an inference from the 1-possibility of $P \& \sim Q$ to its 2-possibility. So we can suppose that the view is correct in order to see what follows.

The Two-Dimensional Argument Against Materialism

On this view, a world may verify P without satisfying P. The secondary intension of P requires that certain specific properties such as mass, spin, and charge are distributed in a certain specific way across space–time, with appropriate causal and nomic relations among them. The primary intension of P requires only that whatever properties play the mass role, the spin role, and charge role are distributed in this way. If W is a world where these roles are played by properties other than mass, spin, and charge (we might say that they are played by ‘schmass’, ‘schmin’, and ‘schmarge’), which are otherwise distributed in the right way over space–time and have appropriate causal and nomic relations among them, then W will verify P, but it will not satisfy P. Here we might say that the physics of W has the same *structural* profile as physics in the actual world, but that it has a different *intrinsic* profile, in that it differs in the intrinsic properties that fill this structure. To verify P, a world must have the right structural profile, while to satisfy P, a world must have the right structural and intrinsic profile.

It follows that premise 3 is not guaranteed to be true. Because a world can verify P without satisfying P, it may be that $P \& \sim Q$ is 1-possible but not 2-possible. However, this requires that P and Q be related in a certain specific way. In particular, it requires that some worlds that verify P also verify $\sim Q$, while no worlds that satisfy P also satisfy $\sim Q$. This requires in turn that some worlds that have the same structural profile as the actual world verify $\sim Q$, while no worlds that have the same structural (p. 322) and intrinsic profiles as the actual world satisfy $\sim Q$. We can assume for the moment that the primary and secondary intensions of Q coincide. Then we can put all this by saying that the falsity of (3) requires that the structural profile of physics in the actual world does not necessitate Q, but that the combined structural and intrinsic profiles of physics in the actual world do necessitate Q.

This idea—that the structural properties of physics in the actual world do not necessitate the existence and/or nature of consciousness, but that the intrinsic properties of physics combined with the structural properties do—corresponds to a familiar view in the metaphysics of consciousness. This is the view that I have elsewhere called *Russellian monism* (or type-F monism, or panprotopsychism). On this view, consciousness is closely tied to the intrinsic properties that serve as the categorical bases of microphysical dispositions. Russell and others held that the nature of these properties is not revealed to us by perception (which reveals only their effects) or by science (which reveals only their relations). But it is coherent to suppose that these properties have a special nature that is tied to consciousness. They might themselves be phenomenal properties, or they might be *proto-phenomenal* properties: properties that collectively constitute phenomenal properties when organized in the appropriate way.

Russellian monism is an important view on the mind–body problem. I think that it is certainly not ruled out by the conceivability argument and by related arguments. If Russellian monism is true, then when we conceive of zombies, we hold fixed the structural properties of physical systems in the actual world, but not their intrinsic properties (which are proto-phenomenal properties). If we consider these intrinsic proto-phenomenal properties to be physical properties, then Russellian monism will qualify as a

The Two-Dimensional Argument Against Materialism

form of physicalism. But because it relies on speculation about the special nature of the fundamental properties in microphysics, it is a highly distinctive form of physicalism that has much in common with property dualism, and that many physicalists will want to reject.¹

In any case, we can now close the loophole in the previous argument as follows:

- (1) $P \& \sim Q$ is conceivable.
- (2) If $P \& \sim Q$ is conceivable, $P \& \sim Q$ is 1-possible.
- (3) If $P \& \sim Q$ is 1-possible, $P \& \sim Q$ is 2-possible or Russellian monism is true.
- (4) If $P \& \sim Q$ is 2-possible, materialism is false.

-
- (5) Materialism is false or Russellian monism is true.

(p. 323) This argument is valid. I discussed the case for premises (1), (2), and (4) earlier, and I have just now argued for premise (3). I think that (5) is the proper conclusion of the conceivability argument. For the reasons given above, such arguments (and also related arguments such as the knowledge argument and the property-dualism argument) cannot exclude Russellian monism, and Russellian monism is arguably a form of physicalism, if a distinctive and radical kind. So the possibility of Russellian monism needs to be explicitly acknowledged as an option in the conclusion.

A couple of minor notes on the argument. First, to be fully explicit the argument might take the truth of Q as a premise. If Q were false, the ground for accepting (4) would collapse. In the less explicit version of the argument above we can consider the truth of Q part of the case for accepting premise (4). In fact, for reasons given earlier, the case for (4) requires that Q is a *positive* truth about consciousness. Alternatively, one can remain silent on whether Q is a positive or negative truth, and handle this matter by conjoining P with a 'that's-all' clause asserting that the world is a minimal world in which P (or, equivalently, by building such a that's-all clause into P).

Second, it is worth noting that (contrary to a common supposition), the assumption that Q has the same primary and secondary intensions is not necessary for the argument for (5) to go through. To see this we can consider the version of the argument where we adjoin a 'that's-all' clause to P . From (1) and (2) we can derive the conclusion that there is a minimal world verifying P in which the primary intension of Q is false. If P has the same primary and secondary intensions, then this world will be a minimal P -world in which the primary intension of Q is false. This world must differ from our world, because the primary intension of Q is true in our world. (There is a small loophole here arising from the possibility that this world differs merely in the location of the centre of the relevant centred world. I discuss this loophole in the expanded version of this paper.) It follows that there is a minimal P -world that is not a duplicate of our world, so that physicalism is false of our world. It could be that strictly speaking physicalism will be true of *consciousness*, because P necessitates Q , but physicalism will be false of properties closely associated with consciousness; namely, those associated with the primary

The Two-Dimensional Argument Against Materialism

intension of Q. We might think of this sort of view as one on which phenomenal properties are physical properties that have non-physical properties as modes of presentation.

Alternatively, if P has different primary and secondary intensions, then by the reasoning given in the earlier discussion of premise (3) one can conclude that either there is a minimal world satisfying P in which the primary intension of Q is false (which again entails the falsity of physicalism), or that the primary intension of Q is necessitated by the structural and intrinsic profiles of physics in our world, but not by the intrinsic profiles alone. This view can be considered another form of Russellian monism, in that the intrinsic properties of physics in our world are crucial for constituting the properties associated with the modes of presentation of consciousness. So if Q has distinct primary and secondary intensions, then one will have to formalize premises (3) and (4) somewhat differently, but the argument for (5) will still work just as well.

(p. 324) 18.4 Objections

For reasons of space I will here pass over objections to premise 1. Six such objections are discussed in the expanded version of this chapter: the objection from analytic functionalism, the objection that zombies are *prima facie* but not ideally conceivable, an objection based on expanding the conception of the physical (Stoljar 2001), the objection that zombies presuppose epiphenomenalism (Perry 2001), the objection based on the fact that zombies appear to make judgements about consciousness (Thomas 1998; Kirk 1999; Lynch 2006), and the objection that zombies are negatively but not positively conceivable (Ashwell 2003; Marcus 2004).

I will also pass over some hard-to-classify objections, including the conditional-concepts objection (Hawthorne 2002; Stalnaker 2002; Braddon-Mitchell 2003), the zombie-parity objection (Balog 1999), the indexical objection (e.g. Ismael 1999; Perry 2001), and objections to two-dimensional semantics (e.g. Soames 2004). Again, these objections are discussed at length in the expanded version of this chapter.

Instead, I will concentrate here on objections to the crucial premise (2). It is this premise that bridges the epistemic and modal domains, and it is this premise and associated principles that have attracted the most in-depth philosophical discussion.

Premise (2) says that if $P \& \sim Q$ is conceivable, $P \& \sim Q$ is 1-possible. This premise can be seen as an instance of the general conceivability-possibility thesis CP:

CP: If S is conceivable, S is 1-possible.

The Two-Dimensional Argument Against Materialism

Here 'conceivability' should be understood as ideal primary conceivability, of either the negative or positive variety (I always take 'ideal primary' as understood from here on). The two versions of the thesis that result are equivalent to theses CP⁻ and CP⁺, discussed earlier. Thesis CP⁻ is equivalent to the claim that if $\sim S$ is not a priori, S is 1-possible. The positive version CP⁺, holding that if S is positively conceivable, S is 1-possible, is somewhat weaker than the negative version, as positive conceivability entails negative conceivability but the reverse is not obviously the case. Much of my discussion will apply equally to both CP⁺ or CP⁻, so I will often just speak of CP, except where the distinction is relevant.

The case for premise (2) largely derives from the case for CP, and from here on I will mostly focus on the general principle rather than the specific premise. Of course, if it turns out that the general principle needs to be restricted to a certain class of sentences to be plausible, then the question will arise as to whether P& \sim Q falls into that class.

Why believe CP? In the first instance, the thesis is plausible because there are no clear counter-examples to it. Principles linking conceivability and possibility have been widely accepted in the history of philosophy, but have more recently been questioned because of various counter-examples, such as the Goldbach case (both Goldbach's conjecture and its negation are conceivable but only one is possible) and especially the Kripke cases ('Hesperus is not Phosphorus' is conceivable but not possible). But CP accommodates these examples straightforwardly, with the idealization (p. 325) accommodating Goldbach cases, and the primary/secondary distinction accommodating Kripke cases. If it handles these cases, then the central sources of resistance to conceivability-possibility principles is undermined. But, of course, there may be other possible sources of resistance.

Before proceeding, it is useful to clarify CP by making clear what a counter-example to it would involve. According to the two-dimensional analysis, ordinary Kripkean a posteriori necessities such as 'Water is H₂O' and 'Hesperus is Phosphorus' have a necessary secondary intension but a contingent primary intension. That is, such statements are 2-necessary but 1-contingent: there are centred possible worlds (a Twin Earth world, or a world with distinct morning and evening stars) that verify their negations. When S is an a posteriori necessity of this sort, with a contingent primary intension, we might say that S is a *weak a posteriori necessity*.

By contrast, we can say that an a posteriori necessity is a *strong a posteriori necessity*, or just a *strong necessity*, iff S has a necessary primary intension. Strong necessities are a posteriori necessities that are verified by all centred metaphysically possible worlds. It is not easy to give examples of strong necessities, as all of Kripke's a posteriori necessities appear to be weak necessities. But I will discuss some putative candidates in what follows.

It is easy to see that CP⁻ is equivalent to the thesis that there are no strong necessities. If S is negatively conceivable but not 1-possible, then $\sim S$ will be a strong necessity. If S is a strong necessity, then $\sim S$ will be negatively conceivable but not 1-possible.

The Two-Dimensional Argument Against Materialism

In so far as CP+ is potentially weaker than CP−, the relationship between CP+ and the thesis that there are no strong necessities is not as clear. Certainly any counter-example to CP+ will yield a strong necessity, but the reverse is not obviously the case. To handle this we might define two classes of strong necessities, according to whether they provide counter-examples to CP+ or merely to CP−. Let us say that a *negative strong necessity* is a statement S such that S is 1-necessary and 2-necessary but \sim S is negatively conceivable. The latter condition is equivalent to the requirement that S is not a priori, so negative strong necessities are equivalent to strong necessities as defined above. A *positive strong necessity* is a statement S such that S is 1-necessary and 2-necessary while \sim S is positively conceivable. Then all positive strong necessities are negative strong necessities, but the reverse is not trivially the case. CP− and CP+ are then equivalent to the theses that there are no negative strong necessities and that there are no positive strong necessities respectively.

What would a strong necessity involve? To get an idea, consider a philosophical view on which it is metaphysically necessary that an omniscient being (e.g. God) exists, but on which it is not a priori that such a being exists. Then according to this view, 'An omniscient being exists' (or O) is an a posteriori necessity. Like all a posteriori necessities, O is 2-necessary, and \sim O is negatively conceivable (and also positively conceivable, if we add the plausible claim that it is positively conceivable that there is no omniscient being). If O were an ordinary a posteriori necessity, then O would be 1-contingent: there would be a metaphysically possible world verifying \sim O. But if there is no omniscient being, then it seems that there is no such world. (p. 326) 'There is an omniscient being' does not seem to have any difference in its primary and secondary intensions, so if a world satisfies O, it verifies O. So given that O is 2-necessary, O is 1-necessary. It follows that if this philosophical view is correct, then O is a strong necessity: it is at least a negative strong necessity, and given the positive conceivability claim above, it is a positive strong necessity.

One could put the matter by saying that there is an epistemically possible *scenario* verifying \sim O, but no metaphysically possible *world* verifying \sim O. Here a scenario can be understood as corresponding to a maximal a priori coherent hypothesis, in the way that worlds correspond to maximal metaphysically possible hypotheses. (I give a formal treatment of scenarios in Chalmers 2004 and (forthcoming b), but here I will leave the notion intuitive. (One might call this sort of scenario a *negative scenario*, since it corresponds to a maximal negatively conceivable hypothesis. One could also define a *positive scenario* so that it corresponds to a maximal positively conceivable hypothesis.) The notion of scenarios is not defined in terms of metaphysical possibility, and in particular it is not assumed that scenarios correspond to metaphysically possible worlds. But nevertheless it is plausible that there is an intimate relationship.

For any a posteriori necessity S there will be a scenario verifying \sim S. For example, as 'Water is H₂O' is not a priori, there will be a scenario verifying 'Water is not H₂O'. That is, there will be some maximal a priori coherent hypothesis H (perhaps involving the assumption that the watery stuff is made of XYZ, and so on) such that if we accept H, we

The Two-Dimensional Argument Against Materialism

should accept 'Water is not H₂O'. For ordinary a posteriori necessities these scenarios will correspond closely to centred metaphysically possible worlds, so that there will be a centred world verifying $\sim S$. For example, 'Water is not H₂O' is verified by a centred XYZ-world, where the individual at the centre is and has always been surrounded by clear, drinkable XYZ in the oceans and lakes. There is little reason to doubt that such a world is metaphysically possible, and there is an intuitive sense in which it qualitatively matches the scenario that we imagine when we entertain the hypothesis that water is not XYZ.

When S is a strong necessity, by contrast, there will be a scenario verifying $\sim S$, but this scenario will correspond to no metaphysically possible world. (When S is a positive strong necessity, there will be a positive scenario verifying $\sim S$; when S is a negative strong necessity, there will be a negative scenario verifying $\sim S$.) For example, given the theist view outlined above, there will be a (negative and positive) scenario verifying 'There is no omniscient being', involving some maximally detailed hypothesis under which there is no such being. But on this view there is no centred world that corresponds to this scenario, and there is no centred world that itself verifies 'There is no omniscient being'. We might put this intuitively by saying that on this view the space of (centred) metaphysically possible worlds is *smaller* than the space of epistemically possible scenarios, at least in the relevant respect. On this view, there are scenarios that correspond to no world.

To bring this back to the mind-body case: take the paradigmatic type-B materialist who holds that premise (1) is true, premise (2) is false, and materialism is true. On this view, the material conditional $P \supset Q$ (which is itself the negation of $P \& \sim Q$) is a strong necessity. The truth of materialism implies that it is 2-necessary, the truth (p. 327) of (1) implies that it is a posteriori and its negation is 1-conceivable, but the falsity of (2) implies that its negation is 1-necessary. On this view, there will be a *scenario* verifying $P \& \sim Q$, including various specific zombie scenarios. But these scenarios will correspond to no metaphysically possible world.

Note that the analogue of CP with scenarios instead of worlds is close to trivial: if S is conceivable in the relevant sense, there will automatically be a scenario verifying S (at least if the notions of a scenario and of verification are unproblematic). Even a type-B materialist and a believer in strong necessities can accept that principle. They must simply deny that all scenarios correspond to worlds. So CP might be seen as equivalent to the thesis that for every scenario there is a corresponding world.

My view is that for every scenario there is a corresponding world, and that there are no strong necessities. In *The Conscious Mind* I gave the following reasons for this: (i) strong necessities cannot be supported by analogy with other a posteriori necessities; (ii) they involve a far more radical sort of a posteriori necessity than Kripke's, requiring a distinction between conceptual and metaphysical possibility at the level of worlds; (iii) they lead to an ad hoc proliferation of modalities; (iv) they raise deep questions of coherence; (v) they will be brute and inexplicable; and (vi) the only motivation to posit a

The Two-Dimensional Argument Against Materialism

strong necessity in the mind-body case is the desire to save materialism. I still accept most of these reasons, but there is more to say.

In the last decade or so numerous objections to CP have been proposed. These objections fall into a number of classes. The first, and largest, involves attempts at exhibiting clear cases of strong necessities. The second involves attempts at explaining how there might be strong necessities in the phenomenal case (if not elsewhere), by analysing the nature of phenomenal concepts. The third involves general philosophical objections. I will discuss the objections in the first class in the section that follows.

18.5 Are there Counter-examples to CP?

I will discuss a number of putative counter-examples to CP in turn. A number of others are discussed in the expanded version of this chapter: objections from essential modes of presentation, from distinct homophonous expressions (Block 2007), from demonstratives (Schiffer 2003), from dancing qualia (Hawthorne 2007), from inscrutable truths, from the deeply contingent a priori (Hawthorne 2004), from disquotational truths (Yablo 2002), and from response-enabled concepts (Yablo 2002).

18.5.1 Kripke Cases.

It is occasionally proposed that some Kripkean a posteriori necessities are in fact strong necessities. In particular, it is sometimes proposed that coextensive names (p. 328) such as 'Cicero' and 'Tully' may have the same primary intension as well as the same secondary intension. If so, 'Cicero is Tully' is a strong necessity (as it is clearly a posteriori).

In response: When we entertain the hypothesis that Cicero is not Tully, this hypothesis corresponds to specific scenarios that we can elaborate. In particular, the relevant scenarios may involve the hypothesis that the causal chains associated with the names 'Cicero' and 'Tully' pick out different historical individuals. A scenario like this certainly corresponds to a centred metaphysically possible world: there are certainly worlds where the causal chains associated with these words function in this way. And it seems clear that if we discovered that our world were such a world we would reject the hypothesis that Cicero is Tully. So such a world seems to verify 'Cicero is not Tully' (although it may not satisfy 'Cicero is not Tully'). Worlds like this can be found for any Kripkean a posteriori necessity S (as Kripke himself pointed out, in effect), and such worlds will always verify $\sim S$.

One might resist the claim that the world in question verifies S by rejecting the claim that there is an *a priori* entailment (for the speaker in question) from a description of the world in question to S. In response, I think that the sort of considerations in Chalmers

The Two-Dimensional Argument Against Materialism

and Jackson (2001) strongly suggest that these entailments are a priori at least in principle. But even if one rejects this claim, there clearly remains *some* distinctive epistemic relation between the world in question and 'Cicero is not Tully': in particular, it remains the case that if one accepts (hypothetically) that the actual world is qualitatively just like the world in question, and reflects on this hypothesis, then one will reject the claim that Cicero is not Tully. (Note that this is quite unlike the situation that the theist thinks obtains in the God case, where there is no world that stands in this inferential relation to 'There is no God'.) So if one rejects a priori entailments, one can use this sort of inferential relation to define primary intensions, and ordinary Kripkean necessities will always have a primary intension that is false at some world.

18.5.2 Ordinary Macroscopic Truths

Another class of examples comes from the suggestion that consciousness is not alone in its failure to be a priori entailed by microphysical truths. Block and Stalnaker (1999) have suggested that many ordinary macroscopic truths, such as 'Water boils at 100 degrees' (or W), are not a priori entailed in this way. If so, then $P \& \sim W$ is at least negatively conceivable. But $P \& W$ is not possible (at least if water and its properties supervene on the microphysical), and it is not clearly 1-possible, either. If it is not, then $P \& \sim W$ is a counter-example to $CP-$.

In response: Chalmers and Jackson (2001) argue that these macroscopic truths are a priori entailed by P , or at least by $PQTI$, a conjunction of full physical, phenomenal, 'that's all', and indexical information. Of course $P \& \sim Q$, $P \& \sim T$, and $P \& \sim I$ may be negatively conceivable, but the latter two are clearly also 1-possible, while the former is precisely the main topic of dispute. So any failures of entailment by P associated (p. 329) with the failure of entailment to Q , T , or I give no further support to the existence of strong necessities. Such support would require at least a truth M such that $PQTI$ does not a priori entail M , and such that $PQTI \supset M$ is 2-necessary and 1-necessary. The arguments in Chalmers and Jackson (2001) suggest that at no ordinary macroscopic truths M are like this.

In any case, even if one rejects the a priori entailment thesis here, these cases will yield at best exceptions to $CP-$ (i.e. negative strong necessities), not exceptions to $CP+$ (i.e. positive strong necessities). Even if W above is like Q in that $P \& \sim W$ and even $PQTI \& \sim W$ is negatively conceivable, $PQTI \& \sim W$ is not positively conceivable: 'water-zombies' are not positively conceivable in the way that zombies are positively conceivable. So this sort of case leaves $CP+$ unthreatened.

18.5.3 Unknowable Mathematical Truths

The Two-Dimensional Argument Against Materialism

Perhaps the most challenging cases for CP– are mathematical truths M such that M is true (and so necessarily true and 1-necessarily true) but not knowable, and so not knowable a priori. If there are such cases, M is a negative strong necessity (though not a positive strong necessity, as $\sim M$ is not positively conceivable). Here one might appeal to unprovable true mathematical sentences, such as those whose existence is entailed by Godel's theorem.

In response: Unprovability in a given system does not entail non-apriority. For example, the consistency of Peano arithmetic is not provable in Peano arithmetic, but is still plausibly knowable a priori. One can make the case (see Chalmers 2002) that all true statements of arithmetic are knowable a priori at least under an idealization (i.e. our failure to know them a priori is due to certain limitations of our cognitive systems). One might worry about higher set theory, such as the continuum hypothesis, but here it is far from clear that such sentences are determinately true or false, and it is also far from clear that they are not knowable a priori under an idealization. So although these cases provide an interesting challenge to CP–, they do not provide clear counter-examples to CP–. And, as in the previous case, there is no challenge to CP+ here.

18.5.4 Laws of Nature

According to some philosophical views, laws of nature are not just naturally necessary but are metaphysically necessary. Shoemaker (1998) holds such a view, and suggests that it may provide a counter-example to CP.

In response: On some varieties of this view, laws of nature follow a version of the Kripkean model. That is, if mass in the actual world obeys certain laws, then nothing in any counterfactual world counts as mass unless it obeys exactly those laws, so any law involving mass will be necessary. This might hold because (p. 330) of the semantics of mass (which require that a counterfactual property have the same nomic role as actual mass in order to qualify as the referent of 'mass'), or it might hold because of the metaphysics of mass (according to which properties such as mass have their nomic profile essentially, as on Shoemaker's view). On these models we need not deny that there are worlds that correspond to the scenario we conceive when we conceive that mass obeys different laws: it is just that such worlds will contain 'schmass' rather than mass. I think that it is implausible that the modal profile of 'mass' and/or the essential properties of mass are this precise (see Fine 2002 and Sidelle 2002 for arguments), but in any case, this model does not provide a counter-example to CP. In this case, a schmass world may verify the hypothesis that the relevant law of nature is false, so laws of nature are not strong necessities.

To yield strong necessities this sort of view must hold that not only are there no worlds where mass obeys different laws, but there are also no related worlds where something else, 'schmass', obeys those different laws. Here the relevant sort of view is one according to which the fundamental properties and laws of all worlds are the

The Two-Dimensional Argument Against Materialism

fundamental properties and laws of our world (and on which these laws are not knowable a priori). In effect, this restricts the space of metaphysically possible worlds to the space of naturally possible worlds. If this view is correct, then a fundamental law will be a strong necessity: there will be no world corresponding to the scenario that we conceive when we conceive what is false.

I think that there are no good reasons to accept this extremely strong view of laws of nature, and that there are good reasons to reject it. The best reasons to take seriously the hypothesis that laws of nature are necessary come from the Kripke and Shoemaker models above. But nothing in these models supports the strong view, or yields a strong necessity. Rather, the CP thesis can itself be taken as a reason to reject the view.

18.5.5 A Necessary God

I have already noted that the existence of strong necessities is entailed by a theist view on which an omniscient being (or an omnipotent being, or a perfect being) exists necessarily but on which the existence of such a being is not knowable a priori. If we say that a god is by definition an omniscient being (or a perfect being, or whatever), then 'A god exists' will be a strong necessity.

In response: I think that theist views of this sort are to be rejected. If the existence of such a god is knowable a priori, then it may exist necessarily. But if it is not, then one should conclude that such a being exists at best contingently. I cannot go over the arguments for believing in a necessary god here, but they all rest on highly contentious premises, and, once again, CP itself provides an argument against these views. The best way to defend the existence of a necessary god is to argue that a world without such a being is not even conceivable.

Even if one believes that the existence of a god provides a strong necessity, it is not clear that this sort of strong necessity undermines the case against materialism. The debate over materialism uses necessity as a criterion of ontological distinctness: the question of whether physical truths necessitate phenomenal truths is relevant (p. 331) precisely in so far as it bears on the question of whether the phenomenal involves nothing ontologically 'over and above' the physical. But a variety of necessity in which the existence of a god is necessary will not be well suited to this role. On such a view, the existence of a god will be necessitated by physical truths, but such a god will presumably nevertheless be ontologically non-physical. So if the only strong necessities are strong necessities of this sort, connecting ontologically distinct existences, they are no help to physicalism. Under this assumption, then, if $P \& \sim Q$ is conceivable, Q will be something over and above the physical, either because it is necessitated by the physical or because it is tied to the physical only by this sort of strong necessity.

Something similar applies to views on which laws of nature are strong necessities. Even on such views laws presumably connect ontologically distinct properties: if there is a fundamental law connecting properties A and B, this will not ground any sort of ontological reduction of one property to the other. Indeed, if this view is correct, then a dualist view with fundamental phenomenal properties and fundamental laws connecting them to the physical will itself be a view on which the phenomenal is necessitated by the physical. So, again, strong necessities of this sort are no help to the physicalist.

18.5.6 Meta-modal Claims

Yablo (1999) adapts the God case to provide an intriguing argument against CP. According to Yablo, it is at least *conceivable* that there is a necessarily existing god. It is also conceivable that there is no necessarily existing god. So if G is 'It is necessary that there is an omniscient being', then both G and \sim G are conceivable. If so, then by CP both G and \sim G are 1-possible. There appears to be no relevant distinction between the primary and secondary intensions of the expressions involved, so it follows that G and \sim G are 2-possible, or (metaphysically) possible *simpliciter*. So it is possible that it is necessary that there is an omniscient being, and it is possible that it is not necessary that there is an omniscient being. But this is a contradiction, at least given S5 as the logic of the metaphysical modality. If it is possible that S is necessary, then S is necessary, so it is not possible that S is not necessary.

In response: One could respond by denying S5, or by finding relevant two-dimensional structure, but I think these moves are unpromising. One could also respond, more promisingly, by making the observation about the ontological relevance of this sort of necessity above. But I think it is best to deny that it is conceivable that there is a necessarily existing god, at least in the relevant sense of conceivability. Perhaps it is *prima facie* negatively conceivable that there is such a being, in that we cannot obviously rule it out a priori, but I do not think it is conceivable in any stronger sense. I can certainly form no clear and distinct conception of such a god (like many, I was suspicious of the idea the moment I heard about it as a student), and continued rational reflection reveals all sorts of problems with the idea. Once one accepts that it is conceivable that there is no god (and this seems like a much stronger intuition, at least to me), this has a strong tendency to undermine the coherence of the hypothesis that a god exists necessarily.

(p. 332)

The problematic issues here arise because of the double modality: we are conceiving not just of non-modal qualitative features of worlds, but also of what is possible or necessary within those worlds. Conceiving of a god (an omnipotent, omniscient, and benevolent being, say) is arguably not too hard; but to conceive in addition that the being exists necessarily, we have to conceive that the space of possible worlds is such that this god exists in each of them, despite the conceivability of a godless world. That is, we have to conceive that CP is itself false. This is what does all the work in the example: if it is conceivable that CP is false, then (by CP!) it is possible that CP is false. CP is surely necessarily true if it is true at all, so it follows from the possible falsity of CP that CP is false.

Another way to respond to this sort of argument is to restrict the conceivability-possibility thesis to claims about the distribution of non-modal properties within worlds, leaving double modals outside its scope. I think this response would be defensible, and not entirely ad hoc. (CP would then apply to worlds considered non-modally, but not to

The Two-Dimensional Argument Against Materialism

'cosmoses' of possible worlds.) But I prefer to hold on to the stronger thesis, by denying that it is conceivable that CP is false. I hold that CP is a priori, although highly non-trivial, like many theses in philosophy (an a priori argument for CP is given in Chalmers 1999, and in the expanded version of this chapter). If this is correct, then CP is not conceivably false on ideal rational reflection, and it is not ideally conceivable that a necessarily existing god exists.

18.5.7 The Conceivability of Materialism

A closely related meta-modal argument (Marton 1998; Sturgeon 2000; Frankish 2007) that is specific to the mind-body domain proceeds as follows. (i) It is at least conceivable that materialism is true about consciousness. So (ii) it is conceivable that $P \supset Q$ is necessary. By CP (and setting aside two-dimensional structure), it follows that (iii) it is possible that $P \supset Q$ is necessary. But from this it follows (using S5) that (iv) $P \supset Q$ is necessary. Using CP and S5 one can equally infer from the fact that (v) it is conceivable that $P \supset Q$ is not necessary to the conclusion that (vi) $P \supset Q$ is not necessary. But (iv) and (vi) are contradictory. So one should reject CP.

My response here parallels the response in the god case. It may be *prima facie* negatively conceivable that materialism is true about consciousness, but it is not obviously conceivable in any stronger sense. Many people have noted that it is very hard to imagine that consciousness is a physical process. I do not think this unimaginability is so obvious that it should be used as a *premise* in an argument against materialism, but, likewise, the imaginability claim cannot be used as a premise either. And if I am right that CP is a priori, then there is an a priori argument that $P \supset Q$ is not necessary, so that it will not even be ideally negatively conceivable that $P \supset Q$ is necessary.

(p. 333) 18.6 Conclusion

None of the attempts above provides a clear counter-example to CP. In most cases I think there are reasonably straightforward independent grounds for rejecting the claim that the cases in question provide strong necessities. Perhaps the most serious challenges come from mathematical cases such as the continuum hypothesis and from meta-modal cases. These are the cases where, in advance of commitment to CP, independent reasoning does not clearly settle whether or not strong necessities are involved. Still, these are cases where the initial situation is unclear, rather than cases where there is a clear counter-example. If one can argue for CP independently, then these cases are not too much of a threat.

It is also worth noting that cases such as these seem to work best as challenges to CP– rather than to CP+, so that CP+, which is all that is required for the argument against materialism, is relatively unthreatened. We have also seen at various points along the way

The Two-Dimensional Argument Against Materialism

that even if one takes certain cases to involve strong necessities, the existence of such strong necessities will still be compatible with modified versions of CP (say, a version involving ontological necessities in the law/god cases, or a version involving non-modal sentences in the meta-modal cases) that will be strong enough for the anti-materialist argument to go through. So where the consideration of counter-examples to CP is concerned, the anti-materialist seems to be on reasonably strong ground.

Materialists can also object to CP in other ways: for example, by giving theoretical arguments for why it is false without relying on counter-examples, or by giving an explanation of why there should be special counter-examples in the case of consciousness. The second strategy has been especially popular, with a number of theorists arguing that special treatment in the case of consciousness is warranted by the special nature of phenomenal concepts (e.g. Loar 1997; Hill and McLaughlin 1999; Papineau 2002). I discuss objections of this sort in the expanded version of this chapter, and also give a theoretical argument for the truth of CP, arguing that it is grounded in constitutive links between the modal and rational domains.²

References

Ashwell, L. (2003), 'Conceivability and Modal Error', MA thesis (University of Auckland).

Balog, K. (1999), 'Conceivability, Possibility, and the Mind-Body Problem', *Philosophical Review*, 108: 497–528.

(p. 334) Block, N. (2007), 'Max Black's Objection to Mind-Body Identity', in T. Alter and S. Walter, (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (Oxford: Oxford University Press), 249–306.

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

Braddon-Mitchell, D. (2003), 'Qualia and Analytical Conditionals', *Journal of Philosophy*, 100: 111–35.

Chalmers, D. J. (1996), *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press).

— (1999), 'Materialism and the Metaphysics of Modality', *Philosophy and Phenomenological Research*, 54: 473–96.

— (2002), 'Does Conceivability Entail Possibility?' in T. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility* (Oxford: Oxford University Press), 145–200.

— (2003), 'The Content and Epistemology of Phenomenal Belief', in Q. Smith and A. Jokic (eds.), *Consciousness: New Philosophical Perspectives* (Oxford: Oxford University Press), 220–72.

The Two-Dimensional Argument Against Materialism

- (2004), 'Epistemic Two-dimensional Semantics', *Philosophical Studies*, 118: 153–226.
- (2005), 'Phenomenal Concepts and the Knowledge Argument', in P. Ludlow, Y. Nagasawa, and D. Stoljar (eds.), *There's Something About Mary* (Cambridge, Mass.: MIT Press), 269–98.
- (forthcoming a), *The Character of Consciousness* (Oxford: Oxford University Press).
- (forthcoming b), The Nature of Epistemic Space, in A. Egan and B. Weatherson (eds.), *Epistemic Modality* (Oxford: Oxford University Press).
- Chalmers, D. J., and Jackson, F. (2001), 'Conceptual Analysis and Reductive Explanation', *Philosophical Review*, 110: 315–61.
- Fine, K. (2002), 'The Varieties of Necessity', in T. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility* (Oxford: Oxford University Press), 253–82.
- Frankish, K. (2007), 'The Anti-zombie Argument', *Philosophical Quarterly*, 57: 650–66.
- Hawthorne, J. (2002), 'Advice to Physicalists', *Philosophical Studies*, 109: 17–52.
- (2004), 'Deeply Contingent A Priori Knowledge', *Philosophy and Phenomenological Research*, 65: 247–69.
- (2007), 'Dancing Qualia and Direct Reference', in T. Alter and S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (Oxford: Oxford University Press), 195–209.
- Hill, C., and McLaughlin, B. (1999), 'There are Fewer Things than are Dreamt Of in Chalmers' Philosophy', *Philosophy and Phenomenological Research*, 59: 445–54.
- Ismael, J. (1999), 'Science and the Phenomenal', *Philosophy of Science*, 66: 351–69.
- Jackson, F. (1998), *From Metaphysics to Ethics* (Oxford: Oxford University Press).
- Kirk, R. (1999), 'Why There Couldn't Be Zombies', *Aristotelian Society Supplement*, 73: 1–16.
- Loar, B. (1997), 'Phenomenal States (Second Version)', in N. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness* (Cambridge, Mass.: MIT Press), 597–616.
- Lynch, M. (2006), 'Zombies and the Case of the Phenomenal Pickpocket', *Synthese*, 149: 37–58.
- Marcus, E. (2004), 'Why Zombies are Inconceivable', *Australasian Journal of Philosophy*, 82: 477–90.
- Marton, P. (1998), 'Zombies vs. Materialists: The Battle for Conceivability', *Southwest Philosophy Review*, 14: 131–8.

The Two-Dimensional Argument Against Materialism

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

Perry, J. (2001), *Knowledge, Possibility, and Consciousness* (Cambridge, Mass.: MIT Press).

(p. 335) Schiffer, S. (2003), *The Things We Mean* (Oxford: Oxford University Press) (sect. on 'two-dimensional semantics and propositional-attitude content').

Shoemaker, S. (1998), 'Causal and Metaphysical Necessity', *Pacific Philosophical Quarterly*, 79: 59–77.

Sidelle, A. (2002), 'On the Metaphysical Contingency of Laws of Nature', in T. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility* (Oxford: Oxford University Press), 309–36.

Soames, S. (2004), *Reference and Description: The Case Against Two-dimensionalism* (Princeton, NJ: Princeton University Press).

Stalnaker, R. (2002), 'What is it Like to be a Zombie?' in T. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility* (Oxford: Oxford University Press), 385–400.

Stoljar, D. (2001), 'The Conceivability Argument and Two Conceptions of the Physical', *Philosophical Perspectives*, 15: 393–413.

Sturgeon, S. (2000), *Matters of Mind: Consciousness, Reason, and Nature* (London: Routledge).

Thomas, N. J. T. (1998), 'Zombie Killer', in S. Hameroff, A. Kaszniak, and A. Scott (eds.), *Toward a Science of Consciousness II* (Cambridge, Mass.: MIT Press), 171–7.

Yablo, S. (1999), 'Concepts and Consciousness', *Philosophy and Phenomenological Research*, 54: 455–63.

— (2002), 'Coulda, woulda, shoulda', in T. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility* (Oxford: Oxford University Press), 441–92. (p. 336)

Notes:

(*) An expanded version of this article is forthcoming in my *The Character of Consciousness* (Oxford University Press), and is also available online at <http://consc.net/papers/2dargument.html>. Some material in this paper is drawn from Chalmers (1999) and (2002).

(1) A related position arises from views on which laws of nature are necessary (e.g. Shoemaker 1998), and on which there is a lawful connection between physical properties and phenomenal properties in our world. Such a view may hold that it is essential to physical properties that they have this nomic profile, so that there is no world satisfying

The Two-Dimensional Argument Against Materialism

$P \& \sim Q$. Some versions of this view will deny (2), and are discussed later in the chapter. But other versions may accept (2), holding that there is a world verifying $P \& \sim Q$, but will hold that it involves distinct 'schmysical' properties that lack this nomic profile. The resulting view resembles Russellian monism in some respects, but differs from the usual form in taking the connection between physical and phenomenal properties to be nomic in the first instance. Because it turns on this nomic connection, this view does not provide any loophole for materialism: at best, it yields a version of property dualism on which the laws of nature connecting physical and phenomenal properties are necessary.

(2) In the expanded version of this chapter I also consider the relation between the arguments I have discussed here and other anti-materialist arguments, including the knowledge argument, the property-dualism argument, Kripke's modal argument, and Descartes's argument from disembodiment.

David J. Chalmers

David J. Chalmers is Professor of Philosophy and ARC Federation Fellow, Director of the Centre for Consciousness, Research School of Social Sciences, Australian National University, Canberra.

