

Modal arguments against materialism

Michael Pelczar

National University of Singapore

Correspondence

Department of Philosophy, National University of Singapore, Singapore 117570.

Email: pelczar.philosophy@gmail.com

Abstract

We review existing strategies for bringing modal intuitions to bear against materialist theories of consciousness, and then propose a new strategy. Unlike existing strategies, which assume that imagination (suitably constrained) is a good guide to modal truth, the strategy proposed here makes no assumptions about the probative value of imagination. However, unlike traditional modal arguments, the argument developed here delivers only the conclusion that we should not believe that materialism is true, not that we should believe that it is false.

1 | INTRODUCTION

The topic of this paper is modal arguments against materialism. Considering how much attention such arguments have received over the years, it's reasonable to wonder whether there's anything new and useful to say about them. So let me tell you what new and (I hope) useful things I plan to say about them here.¹

The most important thing I have to say is that we can bring modal considerations to bear against materialism without taking conceivability as a guide to possibility. In particular, I argue that the most compelling modal argument against materialism leverages on prima facie modal counterexamples to materialism without making any assumptions about the scope or probative value of imagination.²

I also consider a wider variety of modal arguments than commonly discussed in the literature. A subsidiary finding of the paper is that there's no one-size-fits-all response to these arguments.

In §2, I survey existing modal arguments against materialism. They are all conceivability arguments, which rely on the idea that conceivability (or imaginability) is evidence of possibility. In §3, I review the reasons that materialists and their allies have given for denying that our ability to conceive of zombies, phenomenal inverters, and the like—"phenomenal deviants," as I'll call them—supports the claim that such things are possible. Some of those reasons are questionable, but others seem to pose a real threat to conceivability arguments. This sets things up for §4, where I introduce what I call the Good Modal Argument, which leverages on the prima facie possibility of phenomenal deviants without assuming anything about the limits or probative value of imagination. It emerges that the only

way to challenge the Good Modal Argument is by providing a good reason to believe that phenomenal deviants are impossible. In §5, I consider the main reason materialists have given for believing this—an abductive argument for identifying conscious states with brain states—and argue that it rests on a false abductive principle. In §6, I consider the other reason materialists have given—an argument from mental causation—and argue that it relies on a questionable and inadequately supported premise. §7 concludes the paper.

2 | CONCEIVABILITY ARGUMENTS

Conceivability arguments are by far the best-known modal arguments against materialism. Here we consider two of them: a canonical conceivability argument that depends on a Kripkean error theory of imagination (explained below), and an alternative conceivability argument that doesn't depend on a Kripkean error theory.

Canonical conceivability arguments have two premises: (1) that we can, under certain specified circumstances, conceive of phenomenal deviants (beings that duplicate us physically, but lack some or all of our conscious experiences), and, (2) that if we can thus conceive of phenomenal deviants, then phenomenal deviants are metaphysically possible.³

The simplest conceivability argument goes like this: “We can conceive of phenomenal deviants; if we can conceive of X , then X is metaphysically possible; therefore, phenomenal deviants are metaphysically possible.” If sound, this argument refutes materialism, which implies that any being that duplicates an actual human being in all physical respects must have all the conscious experience that that being has.⁴

The problem with the simple conceivability argument is that it seems people do sometimes conceive of impossible things. For example, Hobbes notoriously claimed to have solved the problem of squaring the circle, and therefore presumably conceived of himself as having squared the circle, even though such an achievement is demonstrably impossible. Similarly, back when it was an open question whether water was H_2O or just HO , chemists in the HO camp presumably conceived of water as being HO , even though (unbeknownst to them) it's impossible for water to be anything but H_2O .⁵

The claim that whatever we can conceive of is possible (equivalently: that we can't conceive of anything impossible) is a crude version of what Peter Kung calls the Kripkean Error Theory of Imagination.⁶ In this crude form, the Kripkean error theory says that we can't really conceive of anything that's metaphysically impossible, and when we take ourselves to conceive of something that turns out to be impossible, what we're really conceiving of is something else that *is* possible, and that we mistake for the impossible thing that we erroneously think we are conceiving of. For example, if you think you can conceive of a world that contains water but no H_2O , you aren't really conceiving of that (since it's impossible): really, you're conceiving of a world that contains neither H_2O nor water (these being the same thing), but something else that outwardly resembles water.⁷

The crude version of the Kripkean error theory isn't very plausible, because of cases like Hobbes and the 18th century chemists, but there is a more sophisticated version of the theory. Instead of saying that we can't conceive of anything impossible, one might say that we couldn't conceive of anything impossible *if we had complete logical, mathematical, and microphysical information*.

The idea here is that even though 18th century chemists could conceive of water as HO , they wouldn't have been able to do so had they known all the microphysical facts, since in that case they'd have known that water was one and the same thing as H_2O , and therefore (the idea goes) would have taken any conceivable scenario alleged to be one in which HO is water as really a scenario in which HO outwardly resembles water without being water.

Similarly for Hobbes: had he known all the logical and mathematical facts, he'd have known that squaring the circle was impossible, and therefore (the idea goes) wouldn't have been inclined to describe any conceivable event as "Hobbes squaring the circle," rather than "Hobbes producing an erroneous geometric construction that he mistakes for squaring the circle."

If this is correct, then even though examples like Hobbes and the 18th century chemists show that conceivability doesn't entail possibility, they leave unscathed the claim that we can infer possibility from conceivability in the light of all logical, mathematical, and microphysical information.

Armed with this more sophisticated version of the Kripkean error theory, opponents of materialism offer the following argument, which I'll call the Canonical Conceivability Argument:⁸

- CC1 We could conceive of phenomenal deviants, even if we knew all the logical, mathematical, and microphysical facts.
- CC2 If we could conceive of *X* even if we knew all the logical, mathematical, and microphysical facts, then *X* is possible.
- CC3 So, phenomenal deviants are possible.

Two things prevent this argument from being entirely persuasive.

First, CC2 is controversial. Suppose I believe (wrongly) that Samuel Clemens and Mark Twain are different men, and suppose that I'm daydreaming about a tennis match between them. In the middle of my daydream, I receive the information that Clemens and Twain are actually one and the same. Can I now return to my daydream? It's not obvious that I can't. No doubt the daydream will appear to me in a different light now: I'll no longer think that the scene I'm imagining is metaphysically, let alone historically, possible. But couldn't I still be imagining it for all that? If so, maybe a materialist who identifies conscious states with brain states could still imagine the brain states existing without the conscious states, even if he knew that this was impossible.⁹

Second, materialists are apt to consider CC1 question-begging. In our present state of ignorance about the brain, how do we know that we aren't like the chemists who thought that water was HO, or the philosopher who thought he had squared the circle? True, we can't see how adding to our existing store of physical information could ever prevent us from conceiving of a zombie or phenomenal invert, but neither could pre-Cantorians mathematicians see how adding to their store of mathematical information could ever prevent them from conceiving of a highest cardinality. Maybe, when it comes to consciousness, we are like the pre-Cantorians.¹⁰

Now let's consider a different conceivability argument that—unlike the canonical argument—doesn't rely on a Kripkean error theory of imagination.

The argument I have in mind takes as its point of departure the platitude that things could have been different in many ways: the gravitational constant needn't have had exactly the value it does, life might never have evolved, a Republican could have won the U.S. presidential election of 1940. We all believe that these things are possible. Why?

Not because we can point to actual confirming instances of them: we can't. Here's a suggestion: we believe they're possible because we can conceive of them, *and have no good reason to think that they are impossible*. If this is correct, it leads to a further suggestion: that it's a norm of everyday modal reasoning that the conceivability of a putative state of affairs establishes its possibility, absent any good reason to think that the state of affairs is impossible.

Let's use "undefeated conceivability" to refer to conceivability in the absence of any good reason to think that the conceived-of state of affairs is impossible: you can form an *undefeated conception* of *X* just in case (1) you can conceive of *X*, and, (2) you have no good reason to believe that *X* is metaphysically impossible. Now consider the following alternative conceivability argument:¹¹

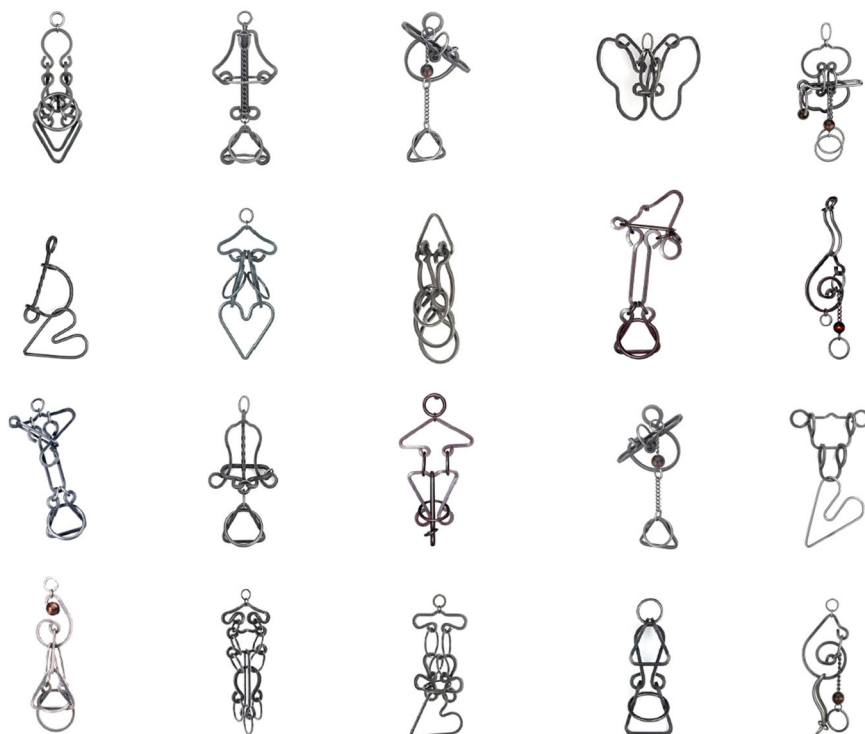


FIGURE 1 Tavern puzzles [Color figure can be viewed at wileyonlinelibrary.com]

AC1 We can form an undefeated conception of a phenomenal deviant.

AC2 If we can form an undefeated conception of *X*, we should believe that *X* is possible.

AC3 Therefore, we should believe that phenomenal deviants are possible.

If sound, this argument wouldn't disprove materialism, but it would show that we should believe that materialism is false, which would serve much the same purpose.

However, the second premise of the argument is false, as the following example shows.

Suppose you're shown a collection of forged iron disentanglement puzzles (a.k.a. tavern puzzles: see Fig. 1). Some are solvable, some not. You know this, but you don't know which ones are the solvable ones (you can't tell just by looking). Call one of the puzzles Puzzle No. 17, and consider the proposition that someone will solve Puzzle No. 17.¹²

You can conceive of someone solving Puzzle No. 17, and you have no good reason to think that a solution is impossible. According to the second premise of the argument above, it follows that you should believe that it's possible that someone will solve Puzzle No. 17. This is wrong.¹³

So, the suggested principle—that undefeated conceivability justifies belief in possibility—turns out to be incorrect. Is there some other principle that succeeds where this one fails?

I don't know. But many materialists argue that we have independent reasons to reject *any* principle purporting to license inferences from the conceivability of phenomenal deviants to their metaphysical possibility. Let's take a look at those arguments now.

3 | IMAGINATION AS A GUIDE TO MODAL TRUTH

The materialists' main complaint about conceivability arguments is that they assume—wrongly, according to materialists—that imagining something under suitable circumstances gives us a good reason to believe that what we imagine is metaphysically possible.¹⁴

In this section, we review the reasons materialists and their allies have given for doubting that imagination is a good guide to possibility. While none of these attempts to downgrade the probative value of imagination clearly succeeds, some of them don't clearly fail either. This is enough to motivate a search for a modal argument that does not rely on claims about the probative value of imagination. I introduce such an argument in §4.

There are two ways you can try to justify skepticism about the probative value of imagination. One is by arguing that our anti-materialist modal intuitions are a cognitive misfire that occurs when we combine different kinds of imagination in certain ways. The other is by arguing that our anti-materialist intuitions are the product of a mental slip by which we mistake mere stipulations for quasi-perceptual revelations. Let's consider these in turn.

Cognitive misfire theories

The classic cognitive misfire theory is Thomas Nagel's. As Nagel describes it, the misfire occurs when we imagine a sensation "sympathetically" (by putting ourselves into a state of mind that resembles the state we'd be in if we had the sensation) and at the same time imagine the sensation's neural correlate (some brain state) "perceptually" (by putting ourselves into a state of mind that resembles the state we'd be in if we observed that brain state). Nagel suggests that this double-act of imagination makes it seem to us that the brain state could exist without the sensation (or vice versa), even though our ability to perform the double-act doesn't make it true or likely that the sensation is anything different from the brain state.¹⁵

According to another influential misfire theory (David Papineau's), our anti-materialist intuitions result from a kind of use-mention fallacy, which Papineau calls the "antipathetic fallacy." Sometimes we think about conscious experience partly by *using* conscious experience similar to the experience we're thinking about, and sometimes we think about conscious experience without using such experience (e.g., by thinking of it as the experience that corresponds to such-and-such a brain state). But the former case is the norm: normally, when we think about an experience, we think of it as *an experience like this one*, where "this one" is an occurrent experience similar to the one we're thinking about (if not that very experience itself). According to Papineau, this tricks us into assuming that the only possible way to conceive of an experience is by means of a similar experience. Since it is possible to conceive of a neural correlate of auditory experience without having any auditory or quasi-auditory phenomenology, we erroneously infer that auditory experiences differ from their neural correlates.¹⁶

Cognitive misfire theories are hampered by the fact that we don't know exactly what psychological mechanism is responsible for our anti-materialist intuitions, or what other beliefs that mechanism produces. Presumably the mechanism isn't a reliable producer of true beliefs, if materialism is true. But since the whole point of cognitive misfire theories is to defend materialism from conceivability arguments, cognitive misfire theorists can't just assume that materialism is true. What they need to provide is *independent* evidence that the psychological mechanisms responsible for our anti-materialist intuitions are unreliable.¹⁷

This task is particularly urgent, considering that there are some seemingly unobjectionable modal beliefs in the same neighborhood as our anti-materialist intuitions.

For example, I believe it's possible for there to be a neuro-phenomenal invert of me: someone whose visual anatomy is wired differently from mine, in a way that results in his having visual experience color-inverted relative to my own. I also believe I could wear goggles that transformed colors into their photographic negatives: a case of optico-phenomenal inversion. Neuro- and optico-phenomenal inverts are possible, as even materialists agree. So, whatever causes us to believe that they are possible causes at least some true beliefs.

These beliefs are very similar to the belief that there could be physically indistinguishable phenomenal inverts. (Arguably, the belief that there could be neurally uninverted phenomenal inverts more closely resembles the belief that there could be neurally inverted phenomenal inverts than it resembles the belief that molecular motion could exist in the absence of heat, or H₂O in the absence of water.) Since it seems reasonable to suppose that these beliefs have similar sources, it also seems reasonable to think that whatever gives us the intuition that phenomenally inverted physical duplicates are possible also gives us at least some correct beliefs. This makes it all the more important for cognitive misfire theorists to provide what they so far have not: examples of uncontroversially *false* beliefs that plausibly have the same source as our intuitions about phenomenal deviants.

A more mundane problem with cognitive misfire theories is that they appear to make false predictions.

For example, suppose I use a combination of perceptual and sympathetic imagination to conceive of Mark Twain: perceptually, I imagine a man with a distinctive moustache and white linen suit; sympathetically, I imagine drinking a mint julep while working out the plot for a story about a boy and an escaped slave. If Nagelian misfire theories are correct, this double-act of imagination should incline me to believe that the white-suited man is different from the man drinking the julep. But it does not.¹⁸

Or consider someone who, like me, can't conceive of a chess position except by visualizing a chess board set up in that position. If the antipathetic theory is correct, this should trick me into thinking that it's impossible to conceive of a chess position except by visualizing a board in that position. But I don't believe that: I'm perfectly aware that one can conceive of a chess position by thinking of its algebraic description, or by imagining it in whatever way a chess player blind from birth imagines chess positions.¹⁹

This is obviously not the last word on cognitive misfire theories, but it seems fair to say that the theories don't receive strong support from the currently available psychological evidence. Let's consider a different objection to taking imagination as a guide to modal truth.

Stipulation theories

Conceivability arguments against materialism assume that the act of imagining a phenomenal deviant does not get its materialism-unfriendly content from a tacit decision by the imaginer to assign such content to the act. If all that distinguishes my mental image of Björk's zombie twin from my mental image of Björk is that the former image, but not the latter, comes with a stipulation that it's an image of a zombie, then my ability to imagine a zombie doesn't count at all towards the claim that zombies are possible. Nobody thinks that stipulation is a guide to possibility.

According to the Stipulation Theory (as I'll call it), when we imagine a phenomenal deviant, like a zombie, our state of mind has no materialism-unfriendly content except what we stipulate for it.²⁰

The stipulation theory relies on a distinction between the phenomenally encoded representational content of a mental image (or act of imagining), and its non-phenomenally-encoded representational content. A mental image's phenomenally encoded representational content—its "phenomenal content," for short—is what the image represents by virtue of having its various phenomenal features (qualia, or

combinations thereof). A mental image's non-phenomenal content is what the image represents other than by virtue of its phenomenal features.²¹

Using this distinction, stipulation theorists argue as follows against taking imagination as a reliable source of modal information (at least as far as that information bears on materialism):

- S1 The phenomenal content of the act of imagining a zombie is identical to the phenomenal content of the act of imagining a normal conscious person.
- S2 The only non-phenomenal content that an act of imagination has is content that we stipulate for it.
- S3 So, the only materialism-unfriendly content of imagining a zombie is content that the imaginer stipulates for the act of imagination.

Call this the Stipulation Argument. If it's sound, our ability to imagine zombies gives us no reason to believe that zombies are possible: how could it, if we imagine zombies only by stipulating that that's what we're doing?

The premises of the Stipulation Argument, though not implausible, are open to challenge. S1 raises questions about what it takes to imagine someone (or something) as conscious or unconscious: does the phenomenal content of my image of my zombie twin really not differ from that of my image of myself? S2 raises questions about the sources of imaginative content: couldn't imagination get some of its representational content in an externalist way (by standing in suitable causal relations to relevant features of the natural or social environment), or maybe by functioning as a *sui generis* cognitive faculty that gives us a kind of quasi-perceptual access to modal reality?

A potentially more serious problem with the Stipulation Argument is that it doesn't seem to work against all of our anti-materialist modal intuitions. For example, if we restate S1 so that it refers to phenomenal inverts instead of zombies, we get something like this:

The phenomenal content of the act of imagining Björk = the phenomenal content of the act of imagining a phenomenal invert of Björk.

The above statement is false. The phenomenal content of imagining a normal person is quite different from that of imagining a phenomenally inverted physical duplicate of that person. When I imagine Björk looking at a lemon and having phenomenally yellow experiences, I do it partly by having yellowish imaginative experience; when I imagine Björk's phenomenal invert having phenomenally blue experiences when looking at the same lemon, I do it partly by having bluish imaginative experience.

Although the phenomenal content of imagining someone with inverted experience differs from that of imagining someone with ordinary experience, one might argue that imagining a phenomenal invert still involves stipulation in a way that undermines the imagining's probative value. For example, Donchadh O'Conaill suggests that I need to stipulate that the inverted experience I imagine belongs to the physical duplicate of Björk that I imagine having the experience.²²

Assessing this suggestion would require going into the question of how we imagine causation: if causal relations are on a par with spatial relations when it comes to how we imagine them, the connection between the imagined inverted experience and the imagined body might be part of the phenomenal content of my act of imagination, rather than something that I merely stipulate. The suggestion that causal relations are part of the phenomenal content of acts of imagination doesn't seem very plausible, though, and even if it's true, there might still be an element of stipulation here: maybe I have to stipulate that the person I imagine with inverted experience has the same physical characteristics as Björk. Whether this would significantly downgrade the probative value of our ability to imagine inverts is a further question that I won't pursue here.

Conceivability arguments rely on claims about the power of imagination to reveal modal truth. Cognitive misfire theories and the stipulation argument raise important questions about those claims. Debate in this area is ongoing, and it's too soon to say who will prevail: the proponents of conceivability arguments or their critics. But I hope this much is clear: a modal argument against materialism that did *not* assume any probative power for imagination would have an important advantage over the arguments we've discussed so far. Let's consider such an argument now.

4 | THE GOOD MODAL ARGUMENT

For the purposes of the following discussion, I define a "good reason" as follows: you have a good reason to believe that p if and only if your evidence justifies you in having a credence in p of greater than half (0.5). Consider now the following naive epistemic principle:

If you have no good reason to believe that p , then you shouldn't believe that p .

A corollary of this principle is that if you have no good reason to believe that X is impossible, then you shouldn't believe that X is impossible. So, if the naive principle is correct, the task of showing that we shouldn't deny the possibility of phenomenal deviants reduces to that of defending the claim that we have no good reason to believe that phenomenal deviants are impossible (no evidence that justifies us in believing that it's more likely than not that phenomenal deviants are impossible).

However, the naive epistemic principle is arguably false, since there are, arguably, beliefs that we're entitled to have without any supporting evidence: "properly basic beliefs," as they're sometimes called. This applies to modal as well as non-modal beliefs. I'm entitled to believe that it's impossible for it to be the case that $1 \neq 1$, but do I have a good reason to believe that this is impossible? I'm not sure I do. The most I can offer is that it just seems obviously impossible for 1 not to equal 1. This doesn't seem like a good reason to believe that it's impossible for 1 not to equal 1 (even if it is a good explanation of why I believe that it's impossible).

Consider then the following, more cautious principle:

We shouldn't believe things that we have no good reason to believe, except in cases where the belief is properly basic.

Call this the Anodyne Principle. It is, I think, about as unobjectionable as an epistemic principle can be without being explicitly tautologous. This is particularly so considering the low standard we've adopted for what counts as a "good reason": you have good reasons to believe that p as long as your reasons entitle you to consider it more likely than not that p . The anodyne principle just says that apart from properly basic beliefs, you shouldn't believe that p , unless your evidence makes it reasonable for you think that p is more likely than not true.

The anodyne principle applies to modal as well as non-modal beliefs. Wendell Willkie lost the 1940 U.S. Presidential election to Franklin D. Roosevelt, but he could have won: nobody thinks that a Willkie victory was metaphysically impossible. Why not? Because our properly basic beliefs don't include the belief that a Willkie victory was impossible, and we have no good reason to think that a Willkie victory was impossible. Anyone who says that a Willkie victory was impossible owes us a good reason to think that it was impossible.

Likewise, anyone who says that phenomenal deviants are impossible owes us a good reason to think that they're impossible. Until someone provides such a reason (or a reason to count as properly basic

the belief that phenomenal deviants are impossible), we shouldn't deny the metaphysical possibility of phenomenal deviants. This suggests the following argument:

- GM1 The belief that phenomenal deviants are impossible isn't properly basic, and we have no good reason to believe that phenomenal deviants are impossible.
- GM2 If the belief that p isn't properly basic, and we have no good reason to believe that p , then we shouldn't believe that p .
- GM3 So, we shouldn't believe that phenomenal deviants are impossible.

I call this the Good Modal Argument. It differs from the others we've considered in two notable respects.

First, it has a different conclusion. The conceivability arguments all concluded either (in effect) that materialism is false, or that we should believe that materialism is false. The Good Modal Argument concludes (in effect) that we shouldn't believe that materialism is true. This is a weaker conclusion than the conceivability arguments', but it's still significant: it means that we shouldn't accept any theory that entails materialism, or reject any theory for conflicting with materialism.

Second, unlike the other modal arguments we've considered, the Good Modal Argument is not a conceivability argument: it doesn't hinge on claims about what we can imagine, or about the probative value of imagination. Where conceivability arguments rely on the claim that imaginability (suitably constrained) is evidence of possibility, the Good Modal Argument relies only on the anodyne principle (that you shouldn't believe what you have no good reason to believe, except when it comes to properly basic beliefs).

The belief that phenomenal deviants are impossible isn't properly basic, as I take it materialists also agree. (Otherwise, why would they waste time arguing for materialism, or defending materialism against prima facie modal counterexamples?) But if the belief that phenomenal deviants are impossible isn't properly basic, and the anodyne principle (GM2) is true, there's only one way to resist the Good Modal Argument: by giving a good reason to believe that phenomenal deviants are impossible.

Materialists have offered two such reasons. Both take the form of arguments for identifying conscious experiences with physical phenomena ("conscious states with brain states," as I'll say for short). According to one, conscious states must be identical with brain states, since otherwise there'd be no plausible way for conscious states to cause physical behavior; we consider this causal argument in §6. According to the other, we have good abductive reasons to identify conscious states with brain states; we consider this argument now.²³

5 | MATERIALISM: THE ABDUCTIVE ARGUMENT

The foremost argument in favor of materialism is that identifying conscious states with brain states is the best way to explain various scientific observations, including observed correlations between conscious states and brain states. Ned Block and Robert Stalnaker explain the basic idea as follows:

Suppose that heat = molecular kinetic energy, pressure = molecular momentum transfer, and boiling = a certain kind of molecular motion... Then we have an account of how heating water produces boiling. If we were to accept mere correlations instead of identities, we would only have an account of how something correlated with heating causes something correlated with boiling. Further, we may wish to know how it is that increasing the molecular kinetic energy of a packet of water causes boiling. Identities allow a

transfer of explanatory and causal force not allowed by mere correlations. Assuming that heat = mke, that pressure = molecular momentum transfer, etc., allows us to explain facts that we could not otherwise explain...

If we believe that heat is correlated with but not identical to molecular kinetic energy, we should regard as legitimate the question of why the correlation exists and what its mechanism is. But once we realize that heat is molecular kinetic energy, questions like this will be seen as wrongheaded.²⁴

By the same token, if we identify conscious states with brain states, we get a simpler and more satisfying account of the phenomena, and quash demands for an explanation of psychophysical correlations. This entitles us to infer that conscious states are indeed identical with brain states—or, so the argument goes.

Let's state the argument more explicitly. Here's a first pass:

If true, the hypothesis that conscious states are identical with brain states explains why consciousness exists (it exists for the same reasons as the brain states), and why conscious states correlate with brain states (they correlate with them because they are them). Any hypothesis that promises explanatory benefits like these is one that we should accept. Therefore, we should believe that conscious states are identical with brain states.

This gets the argument almost but not quite right. The problem is the second premise. The geocentric hypothesis would explain a lot, if it were true, but that doesn't mean we should accept geocentrism. Why not? Because the geocentric hypothesis doesn't cohere with our best available science. The same goes for many discredited or superseded scientific theories. John Wheeler once proposed that the universe contains just a single, time-travelling electron. If true, Wheeler's hypothesis would have explained why all electrons have exactly the same mass ("they" are just a single particle). But scientists don't accept Wheeler's hypothesis, because it doesn't cohere with our best science (Wheeler's hypothesis predicts more positrons than our best science can accommodate).²⁵

Let's restate the argument given above to reflect the importance of coherence with our best science:

- AA1 If true, the hypothesis that conscious states are identical with brain states explains why consciousness exists, and why conscious states correlate with brain states.
- AA2 The hypothesis that conscious states are identical with brain states coheres with our best available science.
- AA3 If hypothesis H has explanatory virtues like the ones mentioned above, and H coheres with our best available science, then we should believe that H is true.
- AA4 So we should believe that conscious states are identical with brain states.

This is the Abductive Argument for materialism.²⁶

The weakest point of the argument is the abductive principle it employs (AA3). There are hypotheses that we do not and should not accept, even though they cohere with our best science and would have explanatory value if true. I'll give two examples.²⁷

In our world, there are two kinds of elementary particles: fermions and bosons. Fermions, of which electrons are the most familiar example, never occupy the same space at the same time. It would be nice to have an explanation for this, since not all particles are mutually excluding in this way: photons, for example, are not (that's what makes lasers possible) and it's not obvious why there should be any particles that differ from photons in this regard.²⁸

Here's a possible explanation: no two fermions occupy the same space at the same time, because at any moment, a fermion just is the region of space it occupies (so that an electron, for example, is identical with the total region of spacetime it occupies: the electron's "worldline"). The reason why we never observe two electrons occupying the same space at the same time is that this would require two regions of space to occupy the same space at the same time, which is impossible.

So, if we identify electrons and other fermions with the regions of space they occupy from moment to moment, we get a simple explanation for why such particles never occupy the same space simultaneously. Furthermore, the hypothesis that each fermion is, at each moment, identical with the spatial region it occupies at that moment—henceforth "the Cartesian fermion hypothesis," or "Cartesian hypothesis," for short—coheres with our best science. At least, it seems compatible with our best physics, which takes no stand on the metaphysical question of whether particles are instantiations of physical properties by regions of spacetime, or instantiations of physical properties by something that occupies those regions.²⁹

Does it follow from all this that we should accept the Cartesian hypothesis? No. If anything, it seems we should reject it for implying that no electron could possibly have had a different spatial location from the one it actually has at any given time. Intuitively, any particle could have had a different worldline from the one it actually has, and the region that any particle actually occupies at a given moment could have contained something else (or nothing) at that moment. If the Cartesian fermion hypothesis were true, none of this would be possible.

Proponents of the Cartesian hypothesis might reply that our intuitions are wrong here, or at least cast in doubt by the fact that they conflict with the simplest possible explanation of the mutually-excluding behavior of fermions (the Cartesian hypothesis). Alternatively, they might concede that the intuitions are correct, but argue for an interpretation of them that's consistent with the Cartesian hypothesis, for example by appealing to David Lewis's modal counterpart theory, according to which "electron *e* could have had a different worldline from the one it actually has" means that there's a possible world in which an electron different from but similar to *e* in certain respects has a different worldline from *e*.³⁰

I don't think these replies are convincing, but the important point for present purposes is that they're controversial, as is therefore the Cartesian hypothesis itself. This is the important point, because the Abductive Argument is premised on the claim (AA3) that we should believe any hypothesis that coheres with our best science and would explain something important if true. It's not enough for the materialist to claim that explanatory promise plus coherence with our best science is a point in favor of an hypothesis. Many hypotheses that we rightly reject have points in their favor (Wheeler's "one electron" hypothesis is one example, the hypothesis that the Sun orbits the Earth is another).

The Cartesian hypothesis would explain something important if true, and it coheres with our best science. These are points in its favor. But they're not sufficient to justify believing that the hypothesis is true. Thus we have a counterexample to AA3.

Here's another interesting fact about our world: its total entropy always increases over time. And here's a possible explanation of this fact: time is identical with a monotonically increasing entropy gradient, and moments of time with entropic states occurring as parts of that gradient. Call this "the entropic hypothesis."

If the entropic hypothesis were true, it would explain why entropy always increases in our universe: it always increases, because later times just are times of higher entropy. The hypothesis would also (if true) answer the ages-old question, "What is time?" So, the entropic hypothesis has considerable explanatory potential. And it's compatible with our best available science, according to which entropy does always increase with time.

Does it follow from all this that we should accept the entropic hypothesis? No. Intuitively, it's possible for there to be a world that begins (like ours) with a Big Bang, and ends with a low-entropy Big

Crunch. Such a world is even compatible with the actual laws of physics, as we currently understand them. But such a world would be metaphysically impossible, if time were just entropy. So this is another counterexample to AA3.

Proponents of the entropic hypothesis might reply that our intuitions are wrong here, or at least cast in doubt by the fact that they conflict with the simplest possible explanation of our world's monotonically increasing entropy gradient (the entropic hypothesis). They might also argue that the explanatory promise of the entropic hypothesis justifies revising our understanding of the laws of physics to reflect the alleged impossibility of a world in which entropy doesn't increase monotonically.

Again I don't think the reply is convincing, but again what matters is that it's controversial, which means that the entropic hypothesis itself is controversial. The important point here is not that the entropic theory is false, but that it's not shown to be true by the fact that it coheres with our best science and would explain certain things if correct. This point stands, as long as the entropic hypothesis depends for its justification on controversial assumptions, like the assumption that it's logically impossible for total entropy to decrease, vary non-monotonically, or remain constant over time.

To make an abductive argument for materialism work, materialists have to replace AA3 with a weaker principle of the form:

If hypothesis H is such that (1) H would have explanatory value if true, (2) H coheres with our best science, and, (3) H satisfies condition X, then we should believe that H is true.

—where *water*= H_2O , *heat*=*MKE*, and *mind*=*brain* satisfy condition X, but *electrons*=*regions* and *time*=*entropy* do not. Since I don't know of any credible candidate for X, I conclude that the Abductive Argument fails to give us a good reason to believe that phenomenal deviants are impossible.³¹

6 | MATERIALISM: THE CAUSAL ARGUMENT

In §4, we found that materialists have only one way to resist the Good Modal Argument: by giving us some good reason to believe that phenomenal deviants (zombies, phenomenal inverters, etc.) are impossible. In §5, we considered the main reason materialists have offered: phenomenal deviants are impossible, because their possibility conflicts with the identity of conscious states with brain states, which we have good abductive reasons to accept. I've argued that we do not have good abductive reasons to accept that identity.

In this section, we consider a different argument for identifying conscious states with brain states, the Causal Argument:³²

CA1 Conscious experiences have physical effects.

CA2 Physical phenomena have only physical causes.

CA3 Therefore, conscious experiences are physical entities (e.g., brain states).

Given the necessity of identity, it follows that phenomenal deviants are impossible.

The argument for CA2 is familiar: if there were physical phenomena with non-physical causes, then either the physical domain would not be causally closed (contrary to our best empirical evidence), or there would be an implausible systematic overdetermination of some physical events (various instances of animal behavior) by physical and non-physical causes. One might challenge the closure claim or the implausibility of the overdetermination claim, but I won't do so here: my focus will be on CA1.³³

It's undeniable that (for example) pains and instances of pain-behavior occur in a pattern suggestive of a cause-effect relation between the pains and the behavior: a "causelike pattern," as I'll call it. So, if observing a causelike pattern of events were enough to create a reasonable presumption in favor of a corresponding causal claim, materialists would be entitled to the first premise of the Causal Argument by default, and the burden would be on the argument's critics to find a problem with it.

However, observing a causelike pattern of events does not create a reasonable presumption in favor of a corresponding causal claim. There's a causelike pattern in the passage from one season to another: Spring, Summer, Fall, Winter, Spring ... This doesn't create a reasonable presumption in favor of the claim that each season causes the next. There's a causelike correlation between the occurrence of reddish bull's-eye shaped skin rashes and the subsequent onset of flu-like symptoms (the progression typical of Lyme disease). This doesn't create a reasonable presumption in favor of the hypothesis that the rashes cause the flu-like symptoms.

If we're ever tempted to say that someone justifiably infers a cause-effect relation from an observed causelike pattern, it's only because we tacitly assume that the person has additional background information relevant to the case. For example, you might think that repeated observations of burning candles are enough to justify the belief that heat causes wax to melt. But imagine that the person observing the candles has no relevant background information whatsoever, perhaps because he has recently popped into existence *ex nihilo*. From his observations, such a person isn't in a position to infer a cause-effect relation between the heat and the melting. For all he knows, it's the brightness of the flame that causes the melting; for all he knows, the heat and the melting are effects of a common underlying cause; for all he knows, the hypothesis that heat melts wax conflicts with the laws of physics. This doesn't mean that it would be unreasonable for him to entertain the hypothesis that heat causes wax to melt, but it does mean that it would be unreasonable for him to accept the hypothesis without further investigation.

Even if there were contexts in which observing a causelike pattern were enough by itself to create a reasonable presumption in favor of a cause-effect relation among the events that make up the pattern, contexts in which people's causal judgments are known to be unreliable would not be among them. This is relevant, because people's judgments about the causal structure of their own mental lives are notoriously unreliable. There is a large body of psychological evidence showing that people are quite fallible in their judgments about the causal relations among their mental states and between their mental states and their behavior. People frequently misidentify their motives, attribute their success at problem-solving tasks to insights they never had, make erroneous judgments about the origins of their beliefs and preferences, and mistakenly take themselves to have control over things over which they do not really have any control.³⁴

This unreliability extends to people's judgments about the effects of pain on behavior, the stock example in discussions of epiphenomenalism. A paradigmatic example of pain behavior is withdrawal from the painful stimulus, as when you yank your hand away from a burning hot object that you've accidentally touched. People commonly judge their pain to cause their avoidance-behavior in such cases, and a standard criticism of epiphenomenalism is that it conflicts with this commonsense judgment.

Yet the judgment is wrong. The pain doesn't cause the hand's withdrawal. We have a two-neuron reflex arc from the nervous periphery (e.g., nociceptors in the hand), through the spinal cord (but not the brain), and back to the arm muscle, which reflexively yanks the hand away from the painful stimulus. The reflex goes off automatically, before the brain enters a state of pain or its neural correlate.³⁵

These psychological findings don't show that our judgments about the causal structure of mind-body relationships are always unreliable, but that was not my purpose in citing them. After pointing out that observing a causelike pattern of events does not in general create a presumption in favor of the hypothesis that some of the events in the pattern cause the others, I considered the suggestion that there might be domains of inquiry in which observing a causelike pattern does create such a presumption,

and that the psychophysical domain might be one of them. However, scientists trying to discover the cause of some physical phenomenon (disease, seasonal change, or whatever) don't think that observing a relevant causelike pattern entitles them to a corresponding causality claim; so, we shouldn't think that observing causelike psychophysical patterns entitles us to claim that experience causes behavior, unless our pre-theoretical judgments about the origins of our behavior are more reliable than scientists' pre-theoretical judgments about the origins of the phenomena they investigate. The psychological findings cited above suggest that this is not the case.

I've argued that the existence of causelike patterns featuring conscious experience and physical behavior isn't enough to establish that conscious experiences cause physical behavior. What would it take to establish that experiences cause behavior?

When there's a doubt about whether the events in a causelike pattern stand in a cause-effect relation, rather than being common effects of a single underlying cause, the standard way to resolve it is by looking for a mechanism that connects the events. If we can identify a linear causal pathway leading from X to Y, that's enough to establish that X causes Y.

Suppose there are two geysers located several hundred feet apart from one another. Whenever Geyser X sends up a spout of hot water, Geyser Y sends up a spout about a minute later. The geysers' activity therefore exists in a causelike pattern. However, some background knowledge about geysers leads us to believe that Geyser X doesn't cause Geyser Y to spout, but that the geysers are connected to the same subterranean aquifer heated by a single geothermal source, which is the common cause of both geysers' spouting.

However, suppose that further investigation reveals that the geysers are actually not supplied by the same aquifer. It turns out that whenever Geyser X spouts, it causes vibrations in the underlying bedrock, which leads to a temporary widening of cracks in the rock, through which water from a nearby pond flows via a narrow cavern or crevice, eventually reaching the hot rock underneath Geyser Y, which then spouts. Once we've identified this mechanism, we conclude that we were wrong about the geysers: the spouting of Geyser X does, after all, cause that of Geyser Y.

Can materialists identify a mechanism by which conscious experiences cause physical behavior? Can they for example identify a mechanism by which pains cause pain behavior?

They can certainly identify mechanisms by which various brain states—the neural correlates of pain—cause pain behavior (or at least some pain behavior). The neuroscience of pain is far enough advanced that we can describe those mechanisms in some detail, and as neuroscience improves, it will no doubt describe the neural mechanisms by which other brain states cause physical behavior. But the question here isn't whether materialists can identify mechanisms by which the neural correlates of pain cause pain behavior: it's whether they can identify mechanisms by which *pain* causes pain behavior.

Here the materialist is at a serious disadvantage. The only mechanisms that materialists can call on to explain the connection between pain and pain behavior are physical mechanisms. Working back from the pain behavior, which is physical, we find that its immediate cause is certain events taking place in the peripheral nervous system. These events in turn arise from events deeper in the nervous system, and so forth until we arrive at the neural correlates of pain (C-fiber stimulation, or whatever). But how does pain get into the picture? It's hard to see how materialists can identify a mechanism by which pain causes pain behavior, except by identifying a mechanism that links the neural correlates of pain with pain behavior, and then identifying those neural correlates with pains.

In short, materialists can't identify a mechanism by which pain causes pain behavior without assuming that pains are physical entities (the brain states that are pains' neural correlates). But this identification is precisely what the Causal Argument was meant to establish.

To sum up: the first premise of the Causal Argument is that conscious experiences have physical effects. The evidence usually cited for this premise—the causelike patterns in which conscious

experiences and various physical behaviors figure, and the testimony of conscious subjects—are either inconclusive or unreliable. To put the premise on a solid basis, one would have to identify some mechanism by which experiences cause physical behavior. But this strategy is unavailable to materialists, since in order to pursue it successfully, they must already have established what the Causal Argument was meant to show: that conscious experiences are brain states.

7 | CONCLUSION

There are *prima facie* possibilities that, if genuine, conflict with materialism. Modal arguments attempt to use this fact against materialism. The best-known modal arguments are conceivability arguments, which rely on the claim that our ability to imagine or conceive of a state of affairs under suitable circumstances (e.g., in the light of complete logical, mathematical, and microphysical information) gives us a good reason to consider that state of affairs metaphysically possible. In §2, we reviewed conceivability arguments, and in §3 we considered objections to the claim that suitably constrained conceivability is evidence of metaphysical possibility. In §4, we put forward a new modal argument against materialism that doesn't rely on the conceivability claim. To resist that argument, materialists must provide a good reason to believe that phenomenal deviants (zombies, phenomenal inverters, etc.) are impossible. In §5, we considered one of the reasons materialists have offered (an abductive argument); in §6, we considered the other reason (a causal argument). We criticized the abductive argument for relying on a suspect abductive principle, and the causal argument for relying on a questionable and under-supported premise.

Our conclusion: we shouldn't believe that phenomenal deviants are impossible, and therefore shouldn't believe that materialism is true.

The most obvious advantage of the Good Modal Argument over conceivability arguments is that it doesn't depend on controversial claims about the limits or probative value of imagination. That doesn't mean that imaginability is a red herring in this debate, but it does mean that materialism faces a threat that's independent of questions about what we can conceive of and what follows from our ability to conceive of it.

A related advantage of the Good Modal Argument is that it focuses the debate on arguments for materialism, unlike conceivability arguments, which focus the debate on questions about the nature, limits, and epistemic power of imagination. Such questions are interesting in their own right, but a debate that centers on them is bound to seem less threatening to materialists than one centered on the question whether there are any good arguments for materialism. The Good Modal Argument takes the battle to the enemy.

Which brings us to the final advantage of the Good Modal Argument. For the past fifty years, the default stance in debates about consciousness has been that the burden is on those who are skeptical about materialism to justify their skepticism, rather than on materialists to justify their materialism. The Good Modal Argument changes that: it shifts the burden back to the materialist side, where some of us think it has always belonged.³⁶

ENDNOTES

¹ By "materialism" I understand any theory according to which everything supervenes on the physical; i.e. (on one common definition of "supervenience") according to which every possible world physically indistinguishable from ours is indistinguishable from ours simpliciter. There are ways to refine this definition, but the purpose of such refinements is to overcome technical difficulties orthogonal to the aims of this paper; see (Chalmers, 1996, 38-41), (Jackson, 1998, 9-14), (Stoljar, 2010, 133-39), and Blumson and Tang (2015).

- ²The best-known *prima facie* modal counterexamples to materialism are zombies and phenomenal inverters. A zombie is a being that has no conscious experience, despite being physically identical to some actual conscious person (like you). An example of a phenomenal inverter is a perfect physical duplicate of you whose visual experience is color-inverted relative to yours (like a photographic negative).
- ³ X is a phenomenal deviant iff X physically duplicates some actual conscious being Y , but fails to have all of the conscious experiences that Y has. Zombies are extreme examples of phenomenal deviants: they duplicate us physically while having no conscious experience at all. Phenomenal inverters are less extreme examples: they have the same bodies and just as much experience as we do, but some experiences we have they don't (like phenomenally yellow experiences of bananas, and phenomenally green experiences of emeralds).
- ⁴Henceforth, by "possible" I mean logically or metaphysically possible, i.e. true or existent in at least one logically or metaphysically possible world. By "necessary" I mean logically or metaphysically necessary, and by "impossible" logically or metaphysically impossible.
- ⁵That water cannot exist except as H_2O has been the dominant view since the appearance of the first edition of Kripke (1980).
- ⁶Kung discusses the Kripkean error theory in Kung (2010), Kung (2016a), and Kung (2016b). The relevant source in Kripke is (Kripke, 1980, 144-54).
- ⁷ X "outwardly resembles" O just in case X has the features by which we distinguish O from other things, prior to knowing what features are metaphysically necessary and sufficient for something's being O . The outward features of water are those by which we pre-theoretically distinguish water from grain alcohol, nitric acid, and other forms of non-water.
- ⁸See (Chalmers, 1996, 93-171) and Chalmers and Jackson (2001).
- ⁹Defenders of the canonical argument might respond that in the case described, I can't really go back to imagining a scene where Clemens and Twain are playing tennis. Alternatively, they might respond by refining the Kripkean Error Theory. I'm not going to explore those lines of response here. For arguments in support of the Kripkean Error Theory, see (Kripke, 1980, 144-54), Levine (1983), Chalmers (2002), (Papineau, 2007, 479), and (Yablo, 2008, 159); for criticism, see Byrne (2007), Fiocco (2007), Kung (2016a), Priest (2016), Berto and Schoonen (2018), and O'Conaill (2019).
- ¹⁰Or maybe we're like Daniel Stoljar's intelligent slugs, who live on a mosaic composed of triangular and wedge-shaped (like this: \triangleright) tiles combined in various geometric patterns, including circular patterns. Due to limitations of their perceptual apparatus, the slugs perceive all of the individual tiles as triangular, even those that are in fact wedge-shaped. Dualist slugs contend that circular features of their mosaic don't supervene on the geometry of the individual tiles and their arrangement in the mosaic, on the grounds that they can conceive of a mosaic that duplicates theirs in its tessellar properties, but contains no circles. The dualist slugs are wrong, of course, and if they knew why—viz., because their mosaic includes certain wedge-shaped tiles that they misperceive as triangular—they'd lose their ability to conceive of a circle-free mosaic that duplicated theirs in terms of its constituent tiles and their arrangement. Analogously, Stoljar suggests, our ability to conceive of zombies might be due to our being "unaware of a type of nonexperiential truth relevant to the nature of experience." (Stoljar, 2006, 3-13, 87ff).
- ¹¹Undefeated conceivability is a form of what Chalmers calls "prima facie conceivability" (Chalmers, 2002, 147).
- ¹²Instead of a collection of tavern puzzles, we could consider a collection of equations concerning which you know that some but not all have integer solutions, without knowing which ones have integer solutions. Then we could consider the proposition that someone will find an integer solution to Equation No. 17, and the argument proceeds from there.
- ¹³We can run the same objection via open mathematical conjectures, like the Goldbach Conjecture: AC2 implies (wrongly) that we should both believe that it's logically possible that the conjecture is true, and believe that it's logically possible that the conjecture is false.
- ¹⁴I construe imagination broadly, so that what's imaginable coincides with what's conceivable, and "we can imagine X " is synonymous with "we can conceive of X ."
- ¹⁵See (Nagel, 1974, 446), and Hill (1997), which elaborates on Nagel's account and extends it to a broader range of anti-materialist intuitions. Andrew Melnyk develops a similar account in Melnyk (2002).

- ¹⁶ See (Papineau, 2002, 161-74). Other examples of this kind of cognitive-misfire argument are found in (Loar, 1990, 90), Tye (1999), and (Perry, 2001, 119-50).
- ¹⁷ The cognitive misfire theorists need to tell a story about our anti-materialist intuitions analogous to the story that psychologists tell about moral beliefs that arise from feelings of disgust. We have good reasons to think that disgust evolved as a mechanism for avoiding certain forms of biological contamination, rather than as a guide to moral truth. Ideally, cognitive misfire theorists should show that our anti-materialist intuitions arise from some psychological mechanism whose evolutionary *raison d'être* is independent of the truth of those intuitions, as the evolutionary *raison d'être* of disgust is independent of the truth of the moral beliefs that disgust engenders.
- ¹⁸ For more examples of this kind, see Doggett and Stoljar (2010).
- ¹⁹ Pär Sundström argues that Papineau's account also incorrectly predicts certain beliefs about phenomenal distinctness: (Sundström, 2008, 141-42).
- ²⁰ See Fiocco (2007), Kung (2010), Kung (2016a), Berto and Schoonen (2018), and O'Conaill (2019).
- ²¹ As is standard in these discussions, I'm using "mental image" broadly, to include all of the phenomenology involved in imagining something, not just visual phenomenology.
- ²² (O'Conaill, 2019, 64-65).
- ²³ By "brain states" I just mean the physical states that serve as the neural correlates of consciousness. A brain state in this sense need not be a state of a biological brain, nor even strictly speaking a state: for example, it might be a process occurring in a supercomputer running a Bostrom-style ancestor simulation, as described in Bostrom (2003). I use "conscious states are identical with brain states" as shorthand for: "for each conscious state, there is a brain state with which that conscious state is identical."
- ²⁴ (Block and Stalnaker, 1999, 23-24).
- ²⁵ See (Feynman, 1998, 163).
- ²⁶ Abductive arguments for materialism go back at least to Place (1956); see also (Smart, 1963, 11-12; 88-105), Lewis (1966), and (Armstrong, 1968, 92-99).
- ²⁷ Another potential problem with the Abductive Argument is that opponents of materialism are likely to reject AA2: according to them, the materialist hypothesis is *not* compatible with our best science, since (in their view) our best science must allow for the metaphysical possibility of phenomenal deviants. The dialectical situation here is murky, however, and I won't pursue this line of criticism in what follows.
- ²⁸ In quantum mechanics, the mutually excluding character of fermions is expressed by the Pauli Exclusion Principle.
- ²⁹ According to Descartes, all material objects are identical with the regions of space they occupy; the Cartesian fermion hypothesis is a variant of Descartes's view, restricted to the case of fermions.
- ³⁰ For Lewis's counterpart theory, see Lewis (1971) and (Lewis, 1986, 192-263). In (Schaffer, 2009, 144-45), Jonathan Schaffer appeals to counterpart theory in defense of the view that material objects are identical with the regions they occupy.
- ³¹ The cases of water and heat differ from those of electrons and time in at least one way: *prima facie* modal counterexamples to the identification of water with H₂O or heat with MKE—but not to the identification of electrons with regions or time with entropy—are always also describable as deceptive non-counterexamples. For instance, instead of describing a hypothetical scenario as one in which MKE exists without heat, you have the option of describing it as a hypothetical scenario in which MKE exists and is heat, but nothing outwardly resembles heat (e.g., nothing gives people hot sensations); similarly for water and H₂O. By contrast, it's hard to see how you could redescribe *prima facie* modal counterexamples to the entropic hypothesis as cases in which entropy increases with respect to something distinct from but outwardly resembling time, or as cases in which something distinct from but outwardly resembling entropy starts decreasing at a certain point. It's also not obvious that we can describe *prima facie* modal counterexamples to the Cartesian hypothesis as cases in which an electron distinct from but outwardly resembling a given electron *e* has a worldline different from *e*'s actual worldline, or in which something distinct from but outwardly resembling the region that *e* actually occupies contains something besides *e*; at least, it's hard to see how to attempt such a redescription without deploying a controversial modal counterpart theory. All of this is cold comfort to the materialist, however, since it's also hard to see how you could describe a hypothetical case in which the brain state corresponding to some

actual pain exists without the pain as a case in which the pain does exist, but lacks the outward features of pain (i.e., painfulness), or as a case in which the brain state does not exist, but something that outwardly resembles it does.

³² I take this to be the central argument of Kirk (1979) and (Papineau, 2002, 17-28).

³³ For a defense of overdeterministic dualism, see Mills (1996).

³⁴ A classic survey of relevant psychological findings is Nisbett and Wilson (1977), which describes a wide range of erroneous causal judgments implicated in various forms of cognitive bias; the seminal study of the illusion of control is Langer (1975). The results reported in these studies have been widely replicated. Such studies also reveal that third-party observers—lay people asked to account for the behavior of experimental subjects—are just as bad as the experimental subjects themselves at identifying the causes of the subjects' behavior; see (Nisbett and Wilson, 1977, 247-48).

³⁵ (Saladin, 2018, 496-97).

³⁶ Thanks to Zach Barnett, Bob Beddor, Ben Blumson, John Burgess, Stuart Derbyshire, Frank Jackson, David Papineau, Abelard Podgorski, Qu Hsueh Ming, David Sanson, and Neil Sinhababu for their comments on previous versions of this paper. Special thanks to an anonymous referee for this journal, whose comments were a model of constructive criticism.

REFERENCES

- Armstrong, D. (1968). *A Materialist Theory of the Mind*. London: Routledge & Kegan Paul.
- Berto, F., & Schoonen, T. (2018). Conceivability and possibility: some dilemmas for Humeans. *Synthese*, 195, 2697–2715.
- Block, N., & Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review*, 108, 1–46.
- Blumson, B., & Tang, W. H. (2015). A note on the definition of physicalism. *Thought: A Journal of Philosophy*, 4, 10–18.
- Bostrom, N. (2003). Are we living in a computer simulation? *Philosophical Quarterly*, 53, 243–255.
- Byrne, A. (2007). Possibility and imagination. *Philosophical Perspectives*, 21, 125–144.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Chalmers, D. (2002). Does conceivability entail possibility? In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and Possibility* (pp. 145–200). Oxford: Clarendon Press.
- Chalmers, D., & Jackson, F. (2001). Conceptual analysis and reductive explanation. *Philosophical Review*, 110, 315–361.
- Doggett, T., & Stoljar, D. (2010). Does Nagel's footnote eleven solve the mind-body problem? *Philosophical Issues*, 20, 125–143.
- Feynman, R. (1998). The development of the space-time view of quantum electrodynamics. In *Nobel Lectures in Physics 1963-1970* (pp. 155–178). Singapore: World Scientific.
- Fiocco, M. O. (2007). Conceivability, imagination and modal knowledge. *Philosophy and Phenomenological Research*, 74, 364–380.
- Hill, C. (1997). Imaginability, conceivability, possibility and the mind-body problem. *Philosophical Studies*, 87, 61–85.
- Jackson, F. (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press.
- Kirk, R. (1979). From physical explicability to full-blooded materialism. *Philosophical Quarterly*, 29, 229–237.
- Kripke, S. (1980). *Naming and Necessity (revised edition)*. Oxford: Basil Blackwell.
- Kung, P. (2010). Imagining as a guide to possibility. *Philosophy and Phenomenological Research*, 81, 620–633.
- Kung, P. (2016a). You really do imagine it: against error theories of imagination. *Noûs*, 50, 90–120.
- Kung, P. (2016b). Imagination and modal epistemology. In A. Kind (Ed.), *The Routledge Handbook of Philosophy of Imagination* (pp. 437–450). Abingdon: Routledge.
- Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32, 311–328.
- Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, 64, 354–361.
- Lewis, D. (1966). An argument for the identity theory. *Journal of Philosophy*, 63, 17–25.
- Lewis, D. (1971). Counterparts of persons and their bodies. *Journal of Philosophy*, 68, 203–211.
- Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.
- Loar, B. (1990). Phenomenal states. *Philosophical Perspectives*, 4, 81–108.
- Melnyk, A. (2002). Papineau on the intuition of distinctness. *SWIF Philosophy of Mind*, 4.

- Mills, E. (1996). Interactionism and overdetermination. *American Philosophical Quarterly*, 33, 105–117.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435–450.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- O’Conaill, D. (2019). What we conceive of when we conceive of zombies. In M. P. Guta (Ed.), *Consciousness and the Ontology of Properties* (pp. 53–66). New York: Routledge.
- Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Clarendon Press.
- Papineau, D. (2007). Kripke’s proof is ad hominem not two-dimensional. *Philosophical Perspectives*, 21, 475–494.
- Perry, J. (2001). *Knowledge, Possibility, and Consciousness*. Cambridge: MIT Press.
- Place, U. (1956). Is consciousness a brain process? *British Journal of Psychology*, 47, 44–50.
- Priest, G. (2016). Thinking the impossible. *Philosophical Studies*, 173, 2649–2662.
- Saladin, K. S. (2018). *Anatomy and Physiology: the Unity of Form and Function* (8th ed.). New York: McGraw Hill.
- Schaffer, J. (2009). Spacetime the one substance. *Philosophical Studies*, 145, 131–148.
- Smart, J. (1963). *Philosophy and Scientific Realism*. London: Routledge & Kegan Paul.
- Stoljar, D. (2006). *Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness*. New York: Oxford University Press.
- Stoljar, D. (2010). *Physicalism*. New York: Routledge.
- Sundström, P. (2008). Is the mystery an illusion? Papineau on the problem of consciousness. *Synthese*, 163, 133–143.
- Tye, M. (1999). Phenomenal consciousness: the explanatory gap as a cognitive illusion. *Mind*, 108, 705–725.
- Yablo, S. (2008). *Thoughts: Papers on Mind, Meaning, and Modality*. Oxford: Oxford University Press.

How to cite this article: Pelczar M. Modal arguments against materialism. *Noûs*. 2021;55:426–444. <https://doi.org/10.1111/nous.12320>