

Why I Am Not A Physicalist

Michael Pelczar

We consider four types of modal or conceivability argument against physicalism: arguments purporting to show that physicalism is false, arguments purporting to show that we should believe that physicalism is false, arguments purporting to show that we should not believe that physicalism is true, and arguments purporting to show that we do not, in fact, believe that physicalism is true. The good news for the physicalist is that arguments of the first type are polemically weak, and arguments of the second type probably unsound. The bad news for the physicalist is that each of the other two types of argument has at least one persuasive instance that the physicalist has no evident way of blocking.

1 Introduction

This paper considers a variety of arguments that aim to put pressure on physicalism. Some of the arguments purport to show that physicalism is false, some that we should believe that physicalism is false, some that we should refrain from believing that physicalism is true, and some that we do, as a plain matter of fact, believe that physicalism is false (or, fail to believe that it's true). The paper proceeds as follows.¹

§2 considers two modal or conceivability arguments that purport to show that physicalism is false. The conclusion of this section is that those arguments are not very successful, at least from a polemical standpoint.

§3 considers modal arguments for the claim that we should believe that physicalism is false. These arguments are suggestive, but, I argue, flawed.

¹I understand physicalism as the view that all the facts about our world metaphysically supervene on the physical facts about it. There are various ways of formulating physicalism more precisely, but the purpose of those refinements is to overcome technical difficulties orthogonal to the arguments of this paper; see (Chalmers, 1996, 38-41), (Jackson, 1998, 9-14), (Stoljar, 2010, 133-39), and (Blumson & Tang, 2015).

§4 considers a modal argument that concludes—not that we should believe that physicalism is false, but—that we shouldn’t believe that physicalism is true. I contend that it’s extremely difficult for a physicalist to resist this argument in good faith.

§5 considers a modal argument that concludes that we do not, in fact, believe that physicalism is true. This is an amended version of an argument that David Papineau attributes to Saul Kripke. I agree with Papineau that the Kripkean argument is sound, but I argue that Papineau underestimates the challenge it poses to would-be physicalists.

All the arguments discussed here are examples of modal or conceivability arguments. However, they do not use the two-dimensional semantic framework that physicalism’s opponents have come to rely on to bolster the modal arguments that currently dominate the debate. A main lesson of this paper is that we can make serious inroads against physicalism, without taking a detour through two-dimensional semantics.

2 Arguments purporting to show that physicalism is false

Modal arguments attempt to deduce the falsity of physicalism from two premises: one to the effect that we can (under certain specified circumstances) conceive of a world that duplicates ours physically, but fails to contain all of the conscious experience that ours contains, and another to the effect that if we can conceive of such a world (under the specified circumstances), such a world is metaphysically possible. The simplest modal argument goes like this:

A1 We can conceive of zombies.

A2 If we can conceive of X , then X is metaphysically possible.

A3 So, zombies are metaphysically possible.

Let’s call this *the basic modal argument* or “Basic Argument,” for short. Most people are willing to grant its first premise. We can conceive of beings that are just like us physically, but who have no conscious experience; for instance, it’s conceivable to me that somewhere in the universe I have an unconscious doppelganger. That’s all that’s meant by saying that we can conceive of zombies.

However, as many have pointed out, the Basic Argument’s second premise is suspect. The problem is that people sometimes conceive of things that are

later discovered to be metaphysically impossible. This happens whenever someone takes himself to have proved a mathematical proposition that turns out to be false. For example, Hobbes notoriously claimed to have solved the problem of squaring the circle; presumably, he conceived of himself as having squared the circle. But such an achievement is demonstrably impossible. Arguably, something similar happens whenever someone misidentifies a natural kind. At one time it was an open question whether water was H_2O or just HO. 18th century chemists in the HO camp presumably conceived of water as being HO. But given (as is nowadays generally accepted) that it's metaphysically necessary that water is H_2O , those chemists conceived of something metaphysically impossible.^{2,3}

The standard response to this criticism is to replace the Basic Argument with a more sophisticated argument, which I'll call *the refined modal argument*, or "Refined Argument" for short:⁴

- B1 We could conceive of zombies, even if we knew all the logical, mathematical, and microphysical facts.
- B2 If we could conceive of X even if we knew all the logical, mathematical, and microphysical facts, then X is metaphysically possible.
- B3 So, zombies are metaphysically possible.

Maybe some 18th century chemists conceived of water as being HO, but they wouldn't have been able to do so if they had known all the microphysical facts, since in that case they'd have known that water consisted of H_2O molecules, and if they'd known that, they'd have no more been able to

²For more cases like this, including the time-travel case cited below, see (Kung, 2010).

³Here and throughout, by "it's metaphysically necessary that water is H_2O ," I mean that it's metaphysically necessary that *if water exists*, then water is H_2O ; obviously, it's not the case that water is H_2O in worlds in which water doesn't exist.

⁴See (Chalmers, 1996, 93-171) and (Chalmers & Jackson, 2001). What follows is just one example of a refined modal argument against physicalism. There are others; for example, there's the inverted-spectrum argument that replaces "world that contained no conscious experience" with "world in which people's visual experience was color-inverted relative to their actual experience." A more significant departure from the basic modal argument that still falls within the broad category of conceivability arguments against physicalism is Philip Goff's transparency argument (Goff, 2017, 106-25). The remarks I'm about to make on the refined modal argument also apply to the transparency argument: it is polemically weak, since a physicalist is apt to insist that in our present state of ignorance about the physical, we're not in a position to say that there is an unbridgeable epistemic gap between the experiential and the purely physical.

conceive of water being something besides H_2O than to conceive of water being something besides water. (Maybe some of the chemists would have *believed* that they could still conceive of a world in which water was HO, mistaking (1) their conception of a world in which stuff that outwardly resembles water is HO for (2) a conception of a world in which water is HO, but that just shows that it's possible to think you can conceive of something that you can't, in fact, conceive of.) Likewise, if Hobbes conceived of himself as having squared the circle, that's only because he didn't know all the logical and mathematical facts; if he had known all the logical and mathematical facts, he'd have known that the statement that he squared the circle entailed a contradiction.

Personally, I think the Refined Argument is sound. But it's polemically weak, since a physicalist is apt to consider its first premise question-begging. How do we know that in our present state of ignorance, we aren't like the chemists who conceived of water as 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 world, but neither could pre-Cantorian mathematicians see how adding to their store of mathematical information could ever prevent them from conceiving of a highest cardinality, and we know how that story goes.^{5,6}

There is one suggestive disanalogy between the chemical and mathematical cases and the case of consciousness, familiar from Kripke.⁷

⁵The point is a familiar one; see, e.g., (Block & Stalnaker, 1999, 13-16).

⁶A physicalist might also try to cast doubt on the refined argument by pointing out that an ancient astronomer who thought that a transcendent immaterial deity was responsible for the the movements of celestial bodies might have used the same style of argument to "prove" that a physical duplicate of our world could be one in which nothing explained planetary motion. The astronomer would reason that even if he knew all the logical, mathematical, and microphysical facts, he could conceive of a physical duplicate of our world that contained no such deity, and consequently no explanation of planetary motion; from this, he would infer (via the principle that conceivability in the light of all logical, mathematical, and microphysical information ensures metaphysical possibility) that planetary motion could be inexplicable in a world physically indistinguishable from ours. Of course, such a world is not possible, as we've known since Kepler. But the antiphysicalist has an easy response to this: unlike the ancient astronomer's deity, conscious experience is not just an explanatory posit, but a datum of (inner) observation. The most that the case of the ancient astronomer shows is that we should apply the relevant principle (that conceivability in the light of all logical, mathematical, and microphysical information entails metaphysical possibility) only in cases in which what's at issue is something that's not just an explanatory posit.

⁷See (Kripke, 1980, 144-54).

Before the HO vs. H₂O controversy was settled, chemists, like everyone else, had a way of distinguishing water from non-water (grain alcohol, nitric acid, etc). There are certain features such that prior to settling the controversy, chemists classified samples as water or non-water accordingly as the samples had or lacked those features. Let's call these the "outward features" of water (although they needn't be evident to the casual or untrained observer), and let's say that anything that has these features "outwardly resembles" water.

An astute 18th century chemist in the water-is-HO camp might have reflected that if, contrary to his hypothesis, water turned out to be H₂O, then what he was tempted to describe as "conceiving of water as HO" might really just be conceiving of a situation in which something that outwardly resembles water is HO, but not a situation in which water is HO. Then he'd have backed off from his claim to conceive of water as HO, and suspended judgement on the question of what, exactly, he was conceiving of.

An astute 21st century philosopher in the consciousness-is-brain-activity camp can't engage in parallel reflections. He can't reflect that if his hypothesis is wrong and consciousness turns out to be something non-physical, then what he's tempted to describe as "conceiving of consciousness as brain-activity" might really just be conceiving of a situation in which something that outwardly resembles consciousness is brain-activity, but not a situation in which consciousness is brain-activity. If he's conceiving of a situation in which something that outwardly resembles consciousness is a brain-activity, then he *is* conceiving of a situation in which consciousness is brain-activity, since anything that outwardly resembles consciousness is consciousness.

After all, for something to bear an outward resemblance to consciousness is just for it to have whatever features prompt us to identify certain things as instances of consciousness, prior to the resolution of the debate over whether consciousness is brain-activity. But those features are phenomenal features—features by virtue of having which a state is such that there's something it's like to be in it (or a process such that there's something it's like to undergo it). It's true that my evidence that an instance of consciousness is present often consists of another person's behavior, but what I take the behavior to be evidence *of* is something that the behavior can identify as an instance of consciousness in the same way that I identify instances of consciousness in me, namely by their possession of phenomenal features.

Suggestive as it is, it's not obvious how the disanalogy between these cases (water versus consciousness) helps proponents of the Refined Argument avoid the charge of question-begging. A physicalist can grant that anything that outwardly resembles consciousness *is* consciousness, but insist that for

all we know, there are complex neural processes that outwardly resemble consciousness, so that detailed knowledge of those processes would prevent us from conceiving of them as occurring in the absence of conscious experience. We can't presently conceive of such processes, but there was a time when people couldn't conceive of sets with too many members to fit onto an endless list. Maybe, when it comes to consciousness, we are like those people.

Or maybe we're like Daniel Stoljar's intelligent slugs, who inhabit a mosaic comprising triangular and slice-of-pie-shaped tiles combined in various geometric patterns, including circular patterns. Due to limitations of their perceptual apparatus, the slugs can detect only triangles and circles. Dualist slugs contend that circular features of the mosaic don't supervene on the arrangement of its fundamental geometric constituents, on the grounds that they can conceive of a mosaic that duplicates theirs in its micro-tessellar properties, but contains no circles. The dualist slugs are wrong, of course, and if they knew why—viz., because their mosaic includes certain non-triangular tiles that they're unable to detect—they'd lose their ability to conceive of a mosaic that duplicated theirs without containing any circles.

Analogously, Stoljar suggests, our ability to conceive of zombies may be due to our being “unaware of a type of nonexperiential truth relevant to the nature of experience.” To the objection that, unlike the slugs, we can't even *imagine* acquiring information that would enable us to conceive of consciousness as a purely physical phenomenon, it may be replied that prior to Gauss, people couldn't imagine acquiring information that would enable them to conceive of space as curved, let alone information that would lead them to conclude that space actually is curved. Yet that is exactly what happened.⁸

3 Arguments purporting to show that we should believe that physicalism is false

The Basic Argument failed, due to known cases in which someone conceived of something that turned out to be impossible. Still, the claim that conceivability lends credence to metaphysical possibility had some initial plausibility, and it's worth considering where this came from.

I suggest it came from everyday reasoning about what's possible. We all believe that there are many ways the world could have been different from how it actually is. The gravitational constant needn't have had exactly the value

⁸For Stoljar's discussion, see (Stoljar, 2006, 3-13) and (Stoljar, 2006, 87ff).

it actually has, life might never have evolved, a Democrat could have won the presidential election of 2016. Why do we think these things are metaphysically possible? Not because we can point to actual confirming instances of them. It looks as though we believe these things are metaphysically possible because we can conceive of them, *and know of no proof that they are metaphysically impossible*.

Whenever we have a good reason to accept something as an example of someone who conceives of a metaphysically impossible state of affairs (or putative state of affairs), the example is a case of someone who conceives of a putative state of affairs that we know to be provably impossible (e.g., describable in a way that reveals a hidden incoherence or contradiction). For instance, since we can derive a contradiction from the proposition that Hobbes squared the circle, or at least have it on good authority that a contradiction is derivable from that proposition, we don't think that Hobbes' ability to conceive of himself as squaring the circle supports the claim that it's metaphysically possible to square the circle. But if we didn't know that squaring the circle was demonstrably impossible, we couldn't point to Hobbes as a counterexample to the claim that conceivability implies metaphysical possibility. Similarly, if you think that our ability to conceive of time travel shows that conceivability doesn't imply metaphysical possibility, that's only because you think that time travel is demonstrably impossible. If you didn't think you knew of any way to prove that it was metaphysically impossible for Doc Brown to travel from 1985 to 1885 (e.g., by deriving a contradiction from the claim that such a voyage took place), you wouldn't take our ability to conceive of Doc Brown travelling from 1985 to 1885 as a counterexample to the claim that conceivability implies metaphysical possibility.

The suggestion here is that there's a reasonable presumption that the conceivability of a putative state of affairs establishes its metaphysical possibility, absent any knowledge of a proof that that state of affairs is metaphysically impossible.

Let's call conceivability in the absence of knowledge of any proof that what's conceived of is metaphysically impossible "undefeated conceivability." The suggestion is that undefeated conceivability justifies belief in metaphysical possibility; call this the Defeat Principle. If it's correct, we can fix the basic modal argument by slightly weakening its conclusion, as follows:⁹

⁹Undefeated conceivability is a form of what Chalmers calls "prima facie conceivability" (Chalmers, 2002, 147). Logically speaking, the conclusion of the argument that follows isn't weaker than the conclusion of the basic argument, since "a zombie world is metaphysically possible" doesn't logically entail "we should believe that a zombie world is

- C1 We can form an undefeated conception of a zombie.
- C2 If we can form an undefeated conception of X, we should believe that X is metaphysically possible.
- C3 So, we should believe that zombies are metaphysically possible.

Call this the *Defeat Argument*. Although it has some initial plausibility, it is subject to the following objection.

According to the Goldbach conjecture, every even number greater than 2 is a sum of two prime numbers. So far, no one has proved or disproved this conjecture. Consequently, I can conceive of the Goldbach conjecture being true; that is, I can imagine it turning out that every even number greater than 2 is a sum of two primes. Furthermore, I know of no contradiction or incoherence in the claim that the Goldbach conjecture is true; if I did, I'd probably be in line for a Fields Medal. According to the principle that undefeated conceivability justifies belief in metaphysical possibility, it follows that I should believe that it's metaphysically possible that the Goldbach conjecture is true.

However, I can also conceive of the Goldbach conjecture being false; that is, I can imagine it turning out that *not* every even number greater than 2 is a sum of two primes. I know of no contradiction or incoherence in the claim that the Goldbach conjecture is false; if I did, I'd be in a position to prove the conjecture (or direct you to a proof of it). According to the Defeat Principle, it follows that I should believe that it's metaphysically possible that the Goldbach conjecture is false.

The upshot is that the Defeat Principle implies that I should believe both that it's metaphysically possible that the Goldbach conjecture is true, and that it's metaphysically possible that the Goldbach conjecture is false. But I know *a priori* that it's not both metaphysically possible that the conjecture is true and metaphysically possible that the conjecture is false. As a mathematical proposition, the conjecture is either necessarily true or necessarily false—true in all metaphysically possible worlds or none.¹⁰

metaphysically possible.” But the latter claim is weaker than the former in the sense that the conjunction of “a zombie world is metaphysically possible” and “we should believe what is true” entails “we should believe that a zombie world is possible,” whereas the conjunction of “we should believe that a zombie world is possible” and “we should believe what is true” doesn't entail “a zombie world is possible.”

¹⁰The use of a mathematical example here is inessential. We could have made the same point with an *a posteriori* necessity, such as the identification of water and H₂O, or Phosphorus and Hesperus.

Let's pause and take stock.

Even though we know of no example of something travelling faster than light, we consider ourselves justified in believing that something could, as a matter of metaphysical possibility, travel faster than light. Why?

Above, we considered the following answer: because we can conceive of something traveling faster than light, and know of no proof that it's metaphysically impossible for something to travel faster than light. The problem with this answer was that when applied to other cases (like the Goldbach conjecture), it wrongly implied that we'd be justified in accepting claims that we know (or, should know) are mutually incompatible. Still, *something* has got to justify our belief that superluminal travel is metaphysically possible, since the belief clearly is justified.

The key question here is: what relevantly distinguishes the proposition that the Goldbach conjecture is true from the proposition that something travels faster than light?

The seemingly obvious answer is that as far as we know, the proposition that something travels faster than light, unlike the proposition that the Goldbach conjecture is true, is neither necessarily true nor necessarily false (metaphysically necessarily, that is). In this respect, the proposition about light also differs from the proposition that the Goldbach conjecture is false, the proposition that water is HO, the proposition that water is H₂O, the proposition that Hesperus is Phosphorus, and the proposition that Hesperus isn't Phosphorus. Unlike these other propositions, we have no good reason to think that the proposition that something travels faster than light is non-contingent. (A proposition is non-contingent just in case it is necessarily true or necessarily false.)

The suggestion is that the "something" that justifies our belief that superluminal travel is metaphysically possible is a combination of two factors: (1) our ability to form an undefeated conception of superluminal travel, and, (2) our lack of any good reason to think that the proposition that something exceeds the speed of light is non-contingent.

The proposed principle here is: if (a) you can conceive of its being the case that p, (b) you know of no proof that it's metaphysically impossible that p, and (c) you have no good reason to think that the proposition that p is non-contingent, then you should believe that it's metaphysically possible that p. Call this the *Amended Defeat Principle*.

We have good reasons to think that "The Goldbach conjecture is true," "The Goldbach conjecture is false," "water is HO," "water is H₂O," "Hesperus is Phosphorus," "Hesperus isn't Phosphorus" are non-contingent. So the Amended Defeat Principle doesn't tell us that we should believe that it's

possible that the Goldbach conjecture is true, or possible that the Goldbach conjecture is false, or possible that water is HO, or possible that water is H₂O, etc., even if we can conceive of these, and even if we know of no proof that they are metaphysically impossible.

But the principle does tell us that we should believe that it's possible for something to travel faster than light, since here clauses (a) through (c) are all satisfied. In particular, the proposition that something travels faster than light satisfies (c), since we have no good reason to think that this proposition is (metaphysically) necessarily true or (metaphysically) necessarily false.

Supposing that we have no good reason to believe that the proposition that there are zombies is non-contingent, we can use the Amended Defeat Principle to show that we should believe that zombies are possible. Unfortunately for the opponents of physicalism, the Amended Defeat Principle is false, for reasons similar to those that undermined the unamended version of the principle.

Suppose that I'm confronted with large collection of forged iron disentanglement puzzles (a.k.a. tavern puzzles). Some are solvable, some not. I know this, but I don't know which ones are the solvable ones (I can't tell just by looking). Call one of the puzzles Puzzle No. 17, and consider the proposition that someone will solve Puzzle No. 17.¹¹

I can conceive of someone solving Puzzle No. 17. I know of no proof that it's metaphysically impossible that someone will solve Puzzle No. 17. And I have no good reason to think that the proposition that someone will solve Puzzle No. 17 is non-contingent—for all I know, Puzzle No. 17 is one of the solvable ones, in which case the proposition that someone will solve it *is* contingent. According to the Amended Defeat Principle, it follows that I should believe that it's metaphysically possible that someone will solve Puzzle No. 17. This is wrong.

Is there some other principle that successfully does the work that the Defeat principles try to do but fail? If so, I don't know what it is. Maybe there is no generic principle that justifies our justified judgements of metaphysical possibility. Maybe particularism is the right view to take, when it comes to the justification of beliefs about unactualized possibilities. If so, there's little chance that we can use modal reasoning to persuade anyone that physicalism is false.

¹¹Instead of a collection of tavern puzzles, we could consider a collection of equations concerning which I 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.

4 An argument that shows that we shouldn't believe that physicalism is true

Even though the Goldbach conjecture and similar examples show that it's false that undefeated conceivability justifies belief in metaphysical possibility, they don't threaten a weaker version of the Defeat Principle. They don't cast doubt on the claim that if you can conceive of its being the case that p , and you don't know of any proof that it's metaphysically impossible that p , then you should *not* believe that p is metaphysically *impossible*. Call this the *Weak Defeat Principle* ("weak," because it gives us advice about when to refrain from believing that something is not possible, rather than advice about when to believe that something is possible).

The Weak Defeat Principle implies that I shouldn't believe that it's metaphysically impossible that the Goldbach conjecture is true, since I can conceive of its being true, and I know of no proof that it's metaphysically impossible for it to be true. The principle also says that I shouldn't believe that it's metaphysically impossible that the Goldbach conjecture is false. In both cases, the principle gets it right: I shouldn't believe either that there are no possible worlds in which the conjecture is true or that there are no possible worlds in which the conjecture is false. I should suspend judgement on the conjecture's possible truth or possible falsity, which is just what the cautious principle recommends.

The Weak Defeat Principle also implies that I shouldn't believe that it's metaphysically impossible that Puzzle No. 17 will be solved, since I can conceive of its being solved, and I know of no proof that its solution is a metaphysical impossibility. Likewise, the Weak Defeat Principle implies that I shouldn't believe that it's metaphysically impossible that the puzzle will not be solved, since I can conceive of its never being solved, and know of no proof that its eventual solution is metaphysically inevitable. Again, the principle gets it right: I should suspend judgement on the question of whether it's metaphysically possible that the puzzle will be solved.

This gives us the following argument against believing physicalism:

- D1 We can form an undefeated conception of a zombie.
- D2 If we can form an undefeated conception of X , we shouldn't believe that X is metaphysically impossible.
- D3 So, we shouldn't believe that zombies are metaphysically impossible.

Call this the *Good Modal Argument*. What it sacrifices in the strength of its conclusion, it gains in irresistibility.

We think that things never exceed the speed of light in vacuo (c), but we don't think that it's metaphysically necessary that this is so. Why not? Because we can conceive of something moving faster than c , and know of no proof that superluminal travel is metaphysically impossible. (Any reason for having the positive belief that superluminal travel *is* metaphysically possible is just this reason plus some additional consideration; more on this below.) If someone says that it's metaphysically impossible for anything to exceed the speed of light, he either has to give us a compelling reason to think that it's metaphysically impossible for anything to exceed the speed of light, or show that there's something special about light that makes the cautious principle inapplicable to cases involving it. Since no one has ever done either, we're justified in refusing to believe that faster-than-light-speed travel is metaphysically impossible.

Likewise, if someone denies the metaphysical possibility of zombies, he either has to prove that the proposition that we have zombie twins is *necessarily* false, or show that there's something special about consciousness that makes the Weak Defeat Principle inapplicable to cases involving it.

The most plausible attempt to show that it's necessarily false that we have zombie twins is a causation-based argument: (1) our conscious experiences have physical effects; (2) physical effects have only physical causes; so, (3) our conscious experiences are physical. Given the necessity of identity, it follows that it's metaphysically impossible for us to have zombie twins.¹²

The weakest link of this argument is (1).

Most non-philosophers do believe (1). If you ask them why, they'll say it's because whenever they're in pain (for example), their bodies act in certain ways. But the existence of such correlations doesn't establish any causal connection between pain and pain behavior; in particular, it doesn't show that pain causes pain behavior. Correlation doesn't imply causation. So, what most people take as evidence for (1) doesn't actually support (1) at all, any more than the correlation between cat populations and plague deaths in medieval European cities showed that cats caused plague.

Is there a better reason to accept (1)? Two reasons have been proposed.

First, there is the argument that we should accept that conscious experiences have physical effects, since that's the simplest explanation of the observed correlation between (e.g.) pain and pain behavior.¹³

¹²See (Kirk, 1979) and (Papineau, 2002, 17-28).

¹³Thus, for example, (Papineau, 2002, 23). The suggestion that it might always be the behavior that causes the pain rather than vice versa is ruled-out on empirical grounds.

Let's grant, for the sake of argument, that the simplest way to account for pain/pain-behavior correlations is by supposing that the pain causes the behavior. Well, there's also a correlation between plant growth and skin cancer: higher (or lower) rates of either come with higher (or lower) rates of both. The simplest explanation for this is that plant growth causes skin cancer (suppose we can exclude the reverse hypothesis, on the grounds that plants evolved before skin). But the simplest explanation is wrong. Plant growth doesn't cause skin cancer: sunshine causes both.

You might object that although the carcinogenic plants hypothesis is the simplest explanation of the correlation between rates of plant growth and rates of skin cancer, it isn't part of the simplest explanation of *everything* under the Sun. That explanation, which takes into account all sorts of phenomena in addition to plant growth and skin cancer, explains the cancer and the verdure as effects of a common cause.

Fair enough, but why think it's any different when it comes to pain and pain-behavior? If you think that the simplest Theory of Everything will count pain behavior as an effect of pain, rather than as an effect (together with pain) of a common physical cause, that's only because you think that the Theory of Everything will count pain as a physical phenomenon. (If you didn't think that the Theory of Everything would count pain as a physical phenomenon, you wouldn't want to say that pain causes pain behavior, since that would commit you to overdetermination or a violation of the causal closure of the physical.) But whether we should count pain as a physical phenomenon is precisely the question that the causal argument was supposed to settle. The simplicity rationale for (1) is therefore question-begging.

A different argument in support of (1) is that if conscious experiences didn't have physical effects, we wouldn't know anything about conscious experiences, or even that there was such a thing as conscious experience. Since we obviously do know that we have conscious experiences, it follows that our experiences do have physical effects; or, so the argument goes.¹⁴

The most careful development of this argument is Robert Kirk's. According to Kirk, there are only three consideration-worthy explanations for how you can know about your own experiences (or at least, about their phenomenal qualities): (i) by your beliefs about the experiences being suitable effects of the experiences, (ii) by your beliefs about the experiences being suitable causes of the experiences, or, (iii) by your beliefs about the experiences being at least partially constituted by the experiences. Kirk argues that options

¹⁴See (Watkins, 1989), (Kirk, 2005, 37-57), and the discussion in (Chalmers, 1996, 172-209).

(ii) and (iii) don't work, leaving us with (i) as the only viable account of our knowledge of consciousness.

It's not clear that an account along the lines of (ii) or (iii) couldn't be made to work, but let's grant that Kirk is right about that. It's also not entirely obvious that case (i) requires experiences to have physical effects; maybe we could make sense of the idea that our beliefs, or at least our justified true beliefs about our own experiences, are non-physical states of some sort. But let's set that aside too.¹⁵

The real problem with Kirk's argument, as I see it, is that it doesn't take into account a fourth way that one could know about one's own experiences.

If you think that a belief has to have an experience among its causes in order to count as knowledge of that experience, why is that? Presumably it's because you think that in order to count as knowledge, a belief that a certain experience occurred (or is occurring) must be a reliable sign or indicator that the experience occurred (or is occurring). After all, one's beliefs, including those that count as knowledge, have among their causes all kinds of things that are irrelevant to the beliefs' contents. If knowledge has to have its object among its causes, that's only because, or only to the extent that, such causation is necessary for reliable indication.

The point that Kirk overlooks is that it isn't necessary.

Take the standard epiphenomenalist picture, in which my belief that I'm in pain has no effects, but does have a cause—some brain event—that also causes my pain. In this case, the occurrence of my belief is as reliable an indicator of the occurrence of the pain as in the case in which the pain causes the belief. Or take my belief that I was in pain yesterday. If the belief counts as knowledge, then, according to epiphenomenalism, it has among its causes a past brain-event that also caused a pain yesterday. In this scenario, the present occurrence of the belief is as reliable an indicator of the past occurrence of the pain as in a case in which the belief counts the pain itself among its causes.

I conclude that the causal argument against the metaphysical possibility of zombies fails, since we've been given no good reason to accept its first premise.

A final argument for the impossibility of zombies appeals directly to considerations of theoretical simplicity.

¹⁵Not setting it aside would mean getting sucked into the dreaded private language argument.

As Ned Block and Robert Stalnaker point out, a major reason for making an empirical identification is that doing so yields better explanations of natural phenomena and an overall simpler world-view:

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.¹⁶

Likewise, if we identify conscious experiences with neural processes, we get a simpler account of the phenomena, and quash demands for an explanation of psychophysical correlations.

Physicalists promoting their view on grounds of simplicity often write as though science always gives simplicity more weight than “mere” prima facie metaphysical possibility. It seems to me that this misrepresents actual scientific thinking.

Our universe began (circa the Big Bang) in a state of extremely low entropy. Statistically, the likelihood of matter being distributed in such a low-entropy way is mind-bogglingly small (Roger Penrose estimates it at 1 in $10^{10^{123}}$). Yet, as far as we know, a different distribution of matter wouldn’t have been nomologically impossible, let alone logically or metaphysically impossible. Everyone agrees, or should agree, that there are metaphysically possible worlds whose initial states aren’t low-entropy states. This is why it

¹⁶(Block & Stalnaker, 1999, 23-24).

seems reasonable to ask for an explanation for why the initial state of *our* world had such very low entropy.¹⁷

Of course, it would simplify things if we equated “being a very low entropy state” with “being the initial state of a physical universe.” Then the question, “Why was entropy so low circa the Big Bang?” would simply not arise. It would be like asking why heat correlates with molecular energy.

Yet, this doesn’t incline us to think that it’s metaphysically impossible for a physical universe to begin in anything but a low-entropy state. So why should the simplification we’d get by equating consciousness with brain processes incline us to think that it’s metaphysically impossible for a physical copy of me to lack consciousness?¹⁸

There’s a mystery about the low-entropy past, and there’s a mystery about consciousness. If high-entropy initial states were metaphysically impossible, the low-entropy past would have a simple explanation, and the first mystery would dissolve. If zombies were metaphysically impossible, the existence of consciousness would have a simple explanation, and the second mystery would dissolve. But we don’t think that the simplifying and mystery-dissolving benefits of denying the metaphysical possibility of high-entropy initial states entitle us to deny the metaphysical possibility of high-entropy initial states. So we shouldn’t think that the simplifying and mystery-dissolving benefits of denying the metaphysical possibility of zombies entitle us to deny the metaphysical possibility of zombies. Ockham’s Razor just doesn’t cut it here.

I said that there were two strategies by which one could try to block the Good Modal Argument: by proving that zombies are metaphysically impossible, or by showing that our intuition that zombies are metaphysically

¹⁷Whether it’s reasonable to expect an answer better than “that’s just the way it is” is another question. For good philosophical discussion of the low-entropy past, see (Price, 1996), (Price, 2004), and (Callender, 2004). For Penrose’s estimate, see (Penrose, 1989, 344).

¹⁸One might suggest that we *should* equate being the initial state of a universe with being a state of low entropy, on the grounds that we should define temporal order in terms of a monotonic entropy gradient. This is objectionable for various reasons, not least of which is that it implies that there’s no temporal order in a universe bookended by low-entropy states. The important point for present purposes is that we don’t reject the metaphysical possibility of such a universe just because that would allow us to accept an entropic analysis of temporal order.

possible is the result of some kind of cognitive illusion. We've considered the first strategy and found it wanting. Now let's consider the second.

Advocates of the second strategy argue that there's something about our psychology that makes us peculiarly error-prone in our modal reasoning about consciousness, in a way that makes the cautious principle inapplicable to cases involving consciousness. I know of three attempts to show that we are prone to such error. They all take the form of debunking arguments: they're all arguments that try to cast doubt on our belief that zombies are metaphysically possible, by showing that it arises from some factor that tends to make us have the belief, regardless of whether zombies really are metaphysically possible.

The first debunking argument comes from Christopher Hill, building on a suggestion from Thomas Nagel.¹⁹

Nagel distinguishes between two kinds of imagination: "perceptual" and "sympathetic."

To imagine something perceptually, we put ourselves in a conscious state resembling the state we would be in if we perceived that thing. To imagine something sympathetically, we put ourselves in a conscious state resembling the thing itself.²⁰

Nagel then suggests that we can debunk our intuition that a conscious mental state could occur without any corresponding physical state, by reference to the mutual independence of the two types of imagination:

When we try to imagine a mental state occurring without its associated brain state, we first sympathetically imagine the occurrence of the mental state: that is, we put ourselves in a state that resembles it mentally. At the same time, we attempt to perceptually imagine the non-occurrence of the associated physical state, by putting ourselves into another state unconnected with the first: one resembling that which we would be in if we perceived the non-occurrence of the physical state. Where the imagination of physical features is perceptual and the imagination of mental features is sympathetic, it appears that we can

¹⁹For Nagel's suggestion, see (Nagel, 1974), and for the main developments thereof, (Hill, 1997) and (Hill & Mclaughlin, 1999).

²⁰(Nagel, 1974, 446).

imagine any experience occurring without its associated brain state and vice versa. The relation between them will appear contingent even if it is in fact necessary, because of the independence of the disparate types of imagination.²¹

A physicalist might try to account for our ability to form an undefeated conception of a zombie in the same way, so as to discourage us from taking this ability as a reason to refrain from believing that a zombie world is metaphysically impossible.

There are two questions, here. First: can we account for our ability to form an undefeated conception of a zombie along the lines Nagel and Hill propose? Second: if so, should that discourage us from taking our ability to form an undefeated conception of a zombie as a reason not to believe that a zombie world is metaphysically impossible?

But we don't have to bother with the second question, since the answer to the first is "No."

Hill suggests that when one conceives of a state of painlessness, one sympathetically imagines an absence of pain. Be that as it may, to conceive of a zombie is to conceive of a being that has no consciousness at all, painful or otherwise. But sympathetically imagining a total absence of experience is ruled-out by the very definition of sympathetic imagination.²²

Another problem with Hill's argument is that it simply fails to persuade. That's an unfortunate feature for any argument to have, but for a debunking argument, it's fatal.

Suppose you demonstrate to me that my belief that my children are exceptionally intelligent is wholly attributable to the fact that I want them to succeed, and fear that they won't succeed unless they're exceptionally intelligent. Suppose you also make it clear (in case it's not already clear to me) that I would have this desire and fear regardless of my children's mental abilities (whatever they may be). Then, as a rational and clear-thinking person, I'll stop believing that my children are exceptionally intelligent; or, if for some reason I can't get rid of the belief, I'll at least be very uncomfortable with it, and do what I can to suspend or disavow it. By the same token, if you tell me something that leaves my belief in the intelligence of my

²¹Ibid.

²²For Hill's remarks on pain (or its absence), see (Hill, 1997, 69-70).

children intact and undisavowed, it follows (assuming that I'm rational and clear-thinking) that you haven't successfully debunked my belief—haven't shown that it arises solely from a factor that's insensitive to my children's intellectual merits (or lack thereof).

The point is a general one. Suppose that a rational and intelligent person, N, believes that p, but that his belief that p is wholly attributable to some factor, X, that's independent of whether it's true or likely that p. If you now reveal to N that this is the situation—i.e., that his belief that p is wholly attributable to something that gives rise to the belief regardless of whether it's true or likely that p—N will stop believing that p, or, if for some reason he can't stop believing that p, he'll come to regard his belief that p as an alien and unwelcome presence in his mind, like a neurotic suspicion that he's being surveilled by the FBI.²³

I doubt that the Nagelian debunking argument has had this effect on anyone, physicalist or anti-physicalist, with respect to the belief that zombies are metaphysically possible. If it has, then why do rational and intelligent people keep on arguing about zombies? Anyway, the Nagelian argument definitely hasn't had this effect on me, from which it follows that my use of two kinds of imagination in my thinking about consciousness (perceptual and sympathetic) isn't what accounts for my belief that zombies are metaphysically possible, or at least not in a way that casts doubt on that belief.²⁴

I don't think I'm setting the bar too high here. It is possible to debunk a modal intuition. Kripke did it with the intuition that water could have been something different from H₂O. All I'm saying is that if a physicalist

²³Andrew Melnyk suggests that our intuition that zombies are possible might be like the Müller-Lyer illusion, in which two lines that we know are equally long appear unequal. (Melnyk, 2002) But in the Müller-Lyer case, I'm not at all inclined to *believe* that the lines are unequal; it's not like I have to keep reminding myself that they're really the same length, on pain of slipping into the belief that they're unequal. Once I put a ruler to the things, I lose any inclination to believe that they differ in length. By contrast, noting that I employ different modes of imagination in my thinking about consciousness—putting a ruler to my imaginative faculties, so to speak—does nothing to diminish my inclination to believe that zombies are possible.

²⁴In Stephen Yablo's terms, the Nagelian debunking argument fails to satisfy "the psychoanalytic standard"; see (Yablo, 2008, 159).

wants to debunk our intuition that zombies are metaphysically possible, he needs to follow Kripke's example.

A different debunking argument also inspired by Nagel's comments comes from David Papineau. Papineau argues that we believe that zombies are metaphysically possible only because we commit a special kind of use-mention fallacy, which he calls the "antipathetic fallacy." Again, the idea is that we have two ways of thinking about conscious experience. Sometimes we think about conscious experience partly by *using* conscious experience, and sometimes we think about conscious experience without using it. (Analogously, I can think about a chess position partly by using a chessboard set up in that position, or I can think of the position without using a chessboard; e.g., by thinking of it in terms of the standard algebraic notation.) Because we sometimes use consciousness to think about consciousness, we are apt to think that we *aren't* thinking about consciousness when we don't use consciousness to think about it, when really we might just be thinking about consciousness in a way that doesn't involve using consciousness—in a way that mentions pain, for example, without using any pain-related phenomenology. Or, so Papineau argues.²⁵

Papineau's argument is subject to the same objections as Hill's.

Suppose we grant that we often use ϕ (or ϕ -related) experience to conceive of experience with ϕ phenomenology. How does that shed light on what happens when we conceive of a total absence of experience, as we do when conceiving of zombies? In this case, there's no relevant phenomenology for us to use to conceive of the contemplated state of affairs.

In any event, Papineau's argument fails to shake our belief in the metaphysical possibility of zombies, with the fatal results already mentioned in connection with Hill's argument. If I believe that p only because I commit some use-mention fallacy, and you point this out to me, I'll stop believing that p. This is particularly so, considering that use-mention fallacies are generally hard to fall for and easy to avoid once pointed out. Since I'm still comfortably believing that zombies are metaphysically possible, I have to

²⁵See (Papineau, 2002, 161-74). This style of debunking argument is pretty popular, with versions of it espoused by (Loar, 1990, 90), (Tye, 1999), and (Perry, 2001, 119-50), among others. I focus on Papineau, since his version of the argument is the most detailed.

conclude that Papineau has failed to show that I'm believing it only because I'm committing a use-mention fallacy.²⁶

The third debunking argument to consider comes from Andrew Melnyk, who picks up on a suggestion discarded by Papineau. The suggestion is that believing a (non-trivial) identity claim requires combining two mental representations in a way that involves something like “mental file merging,” where such merging is psychologically impossible, when one of the representations is phenomenal, and the other non-phenomenal.²⁷

Why are phenomenal representations supposed to be psychologically unmergeable with non-phenomenal (material) ones?

The reason I have in mind is that one kind of phenomenal concept seems to be usable only to refer to a phenomenal state as one undergoes it (“*That* is going on in me now”), and not to be usable to *re-identify* a phenomenal state, not even to re-identify it as *one of those again*. Now if phenomenal concepts of this kind exist, and if concepts in general can be viewed as analogous to files, then a phenomenal concept of this kind will constitute a file that is only temporary, a file that persists only as long as one is undergoing the experience it picks out. But any file corresponding to a material concept will presumably be permanent; at the very least it will permit the re-identification of whatever it picks out. And, on the not too implausible assumption that no temporary file can be merged with a permanent file, it follows that no phenomenal concept of the kind in question can be merged with a material concept, and hence, if believing identity claims is a matter of mental file-merging, that no identity claim framed using a

²⁶The chess analogy points to another problem. If Papineau’s explanation of my intuition that zombies are metaphysically possible is correct, why don’t I also intuit that it’s possible for a chess board to be set up in a certain position in a world in which nothing satisfies the corresponding algebraic description? Pär Sundström argues that Papineau’s account even predicts intuitions of *phenomenal* distinctness where we have none; see (Sundström, 2008, 141-42).

²⁷(Melnyk, 2002), citing (Papineau, 2002, 165). A phenomenal representation of an experience—or “phenomenal concept,” in Papineau’s terminology—is one that uses that experience, or phenomenology related to it, to represent the experience. Non-phenomenal concepts are what Papineau calls “material concepts.”

phenomenal concept of the kind in question and a material concept can be believed.²⁸

One question this raises is why a temporary file couldn't merge with a permanent one. Why can't I use a material concept of pain to think about pain while I'm in pain, and also thinking about it as *this* unpleasant experience? (As in, "*this* pain is accompanying such-and-such brain activity.") Maybe the phenomenal representation of pain isn't as enduring as the material representation, but so what? I can merge two computer files, even if one of them but not the other is infected with a virus that will soon delete the file.

The more serious problems with Melnyk's proposal are the same as those that pertained to the other debunking arguments we've considered. Even if believing a statement of the form, "*This* sort of experience = such-and-such neural process" would require a psychologically impossible mental file merger, how does that explain our ability to believe a statement such as: "There could be a non-conscious being physically just like me"? And—most important of all—if Melnyk's explanation of why zombies seem metaphysically possible to us is correct, why don't we lose our inclination to believe that they're possible when we reflect on Melnyk's proposal?

Suppose I draw plans for a complicated machine containing a large number of gears, cams, levers, push-rods, etc. The best way to establish the design's workability would be to build an actual machine according to the plans, but maybe that's prohibitively expensive or beyond the capabilities of existing manufacturing technology (like Charles Babbage's Analytical Engine). However, I've taken great care in drawing the plans, and after studying them carefully, the design seems workable—I don't see anything mechanically impossible about it (like the gear assembly in Fig. 1). I show the design to my engineer friends and they see no mechanical impossibility in it either. Finally, I consult a panel of clinical psychologists, asking them whether they can think of any perceptual or cognitive illusion that might have blinded me or the engineers to a mechanical impossibility lurking in my design. They know of none. (One of the psychologists notes that a moiré pattern can give rise to optical illusions, but I point out to him that my plans contain no moiré pattern.)

²⁸(Melnyk, 2002).

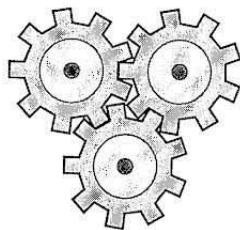


Figure 1: Unworkable gearing

Given all this, am I entitled to judge that my design is mechanically possible? Maybe that's debatable, but I'm certainly *not* entitled to think that the design is mechanically *impossible*. Rather, I'm entitled to find fault with anyone who asserts that the design is mechanically impossible, without producing some heretofore unknown proof of its mechanical impossibility, or pointing out some heretofore unnoticed optical illusion in my plans.

I submit that the situation is exactly the same when it comes to the metaphysical possibility of zombies. If you want to say that the physical facts of our world metaphysically entail the mental facts about it—for example, if you want to advocate physicalism—you better be prepared to back your statement up with some heretofore unknown demonstration that a zombie world is metaphysically impossible, or a debunking argument that actually makes people stop believing that zombies are possible (or at least makes them try to stop believing it). If you can't do either of those things, you shouldn't be a physicalist.

5 Arguments purporting to show that we believe that physicalism is false, or that we don't believe that physicalism is true

The last argument I want to consider is one that David Papineau reconstructs from Saul Kripke's well-known discussion of the identity theory. According to Papineau,

Kripke's challenge isn't to explain how mind-brain identities are a posteriori—as it were, to explain how they can appear possibly false to

people who don't yet believe them. Rather his challenge is to explain why they *still* appear possibly false, even to people who *do* believe them.²⁹

In our discussion of the refined modal argument against physicalism (in §2), we raised the point, due to Kripke, that once it was settled that water was H₂O, erstwhile proponents of the water-is-HO hypothesis would have realized that what they formerly took to be conceiving of water as HO was really just conceiving of a situation in which something that outwardly resembled water is HO. Similarly, once you become convinced that Hesperus is Phosphorus (or Cicero Tully, or whatever), you no longer consider yourself able to conceive of a situation in which Hesperus isn't Phosphorus. Now that you're convinced that Phosphorus is the same thing as Hesperus, you simply have no way to conceive of a situation in which Phosphorus isn't the same thing as Hesperus, since that would require you to conceive of a situation in which a thing was different from itself.

By the same token, if you were convinced that one of your pains was the same thing as some brain state (say, a stimulation of your C-fibers), then you'd have no way to conceive of a situation in which the pain wasn't C-fiber stimulation (e.g., a zombie world). You would look upon anything that you might once have considered conceiving of such a situation as really just a case of conceiving of a situation in which something outwardly resembling pain isn't C-fiber stimulation. However, as Kripke notes, anything that outwardly resembles pain *is* pain.

The upshot is that *if you think that you can conceive of a world in which you lack consciousness, despite having all the physical features you actually have, then you don't believe that consciousness is a brain process*. This yields the following argument, which I'll call the *omissive Kripkean argument*:

E1 If you think that you can conceive of a world in which you lack consciousness, despite having all the physical features you actually have, then you don't believe that consciousness is a brain process.

E2 You do think that you can conceive of a world in which you lack consciousness, despite having all the physical features you actually have.

²⁹(Papineau, 2007, 479). For an earlier construal of Kripke's challenge along the same lines, see (Levine, 1983).

E3 So, you don't believe that consciousness is a brain process.

As I've presented it, the Kripkean argument concludes that you don't believe that consciousness is a brain process. In Papineau's version of the argument, the conclusion is that you *do* believe that consciousness is *not* a brain process; let's call this the *commissive Kripkean argument*.³⁰

The commissive argument replaces E1 with the premise that if you think that you can conceive of a world in which you lack consciousness, despite having all the physical features you actually have, then you believe that consciousness isn't a brain process. Against this, a physicalist might concede that he can conceive of a such world, but deny that this compels him to believe that consciousness isn't a brain process, on the grounds that there are cases in which people have taken themselves to conceive of things that turned out to be metaphysically impossible. Still, if a physicalist (or anyone else) thinks he can conceive of himself lacking consciousness despite having all the physical features he actually has, then he should at least refrain from believing that consciousness *is* a brain process. Otherwise, he'd have to credit himself with the power to conceive of someone as both lacking and having consciousness. So, even if the commissive version of Kripke's argument is open to doubt, the omissive version stands.

Papineau, himself a physicalist, accepts the commissive argument, but points out that its conclusion is consistent with the claim that some of us believe that zombies and suchlike are impossible. Physicalists just have inconsistent beliefs: they believe that zombies are possible, and they believe that zombies aren't possible. But (according to Papineau) that makes the situation sound worse than it necessarily is. At a theoretical level—when thinking about the scientific advantages of physicalism—we're inclined to deny that zombies are possible, while at an intuitive level—when not thinking about the scientific advantages of physicalism—we're inclined to believe that zombies are possible. The question is which level is more important. A physicalist will presumably say it's the theoretical level.

If this response works for the commissive argument, it works for the omissive argument too. On a theoretical level, we believe that zombies are impossible; on an intuitive level, we don't. If the theoretical level is more important, a physicalist might learn to live with this tension.

³⁰See (Papineau, 2007, 478-80).

But is the theoretical level more important than the intuitive level? Not always (recall the low-entropy past). So why think it's more important in this particular case? If you say it's because modal intuitions about consciousness are unreliable, you owe us a good debunking argument.

There's another way a physicalist might try to accommodate the commissive argument.

Belief isn't, or needn't be, all-or-nothing; it comes in degrees. If we look at the situation in terms of graded beliefs ("credences") instead of all-or-nothing ("binary") beliefs, we can say, on the physicalist's behalf, that the commissive argument shows only that physicalists don't assign as high a credence to physicalism as they might like to. That doesn't mean that they don't assign higher credence to the proposition that physicalism is true than to the proposition that it's false.

An immediate problem with this credence-splitting strategy is that it doesn't apply to the omissive version of Kripke's argument, which concludes not that you *have* the belief that zombies are possible, but that you *lack* the belief that zombies *aren't* possible. Here, there's no question of dividing your credences between two propositions—that zombies are possible, and that they aren't—since (if the argument is sound) you don't believe the second proposition at all.

A physicalist might say that the conclusion of the omissive argument should really just be that we don't have a very high (greater than 0.8?) credence that consciousness is a brain process. That is, he might insist on replacing E1 with the premise that if we think we can conceive of a zombie world, then we don't believe *to a very high degree* that consciousness is a brain process. That would leave open the possibility that we believe to a degree greater than 0.5 that consciousness is a brain process.

Suppose that, like most people who have an opinion about such matters, you think you can conceive of a situation in which William Shakespeare exists but Christopher Marlowe doesn't. You've considered the possibility that all you can really conceive of is a situation in which someone who outwardly resembles Shakespeare exists in a Marloweless world, but you are confident that that's not what you're doing. You've reviewed all the evidence that people have offered in support of the theory that Shakespeare and Marlowe were actually the same man (e.g., putative evidence that the works commonly

attributed to Shakespeare were actually penned by Marlowe), and found that none of it holds up to scrutiny.

Where does that put your credence that Shakespeare was Marlowe? Above 0.5? Presumably not. But then presumably your credence that consciousness is a brain process isn't above 0.5 either, if you think you can conceive of a situation in which the brain process exists but consciousness doesn't.

Conclusion

We've considered a variety of broadly anti-physicalist arguments, including arguments that purport to show that physicalism is false, arguments that purport to show that we should believe that physicalism is false, arguments that purport to show that we should not believe that physicalism is true, and arguments that purport to show that we do not, in fact, believe that physicalism is true.

The good news for the physicalist is that arguments of the first two types don't pose a clearly lethal threat to his position. The bad news is that each of the other two types of argument exists in at least one version that the physicalist has no evident way of blocking. The successful arguments—the Good Modal Argument, and the omissive Kripkean argument—are hostage to fortune, in that their success depends on nobody's proving that zombies are impossible, or successfully debunking our intuition that they are possible. I've argued that the best attempts at such a proof or debunking have failed. Until and unless someone comes up with a better response to these arguments, no one should be a physicalist.

References

- Block, Ned, & Stalnaker, Robert. 1999. Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review*, **108**(1), 1–46.
- Blumson, Ben, & Tang, Weng Hong. 2015. A Note on the Definition of Physicalism. *Thought: A Journal of Philosophy*, **4**(1), 10–18.
- Callender, Craig. 2004. There is no puzzle about the low-entropy past. *Pages 240–255 of: Hitchcock, Christopher (ed), Contemporary Debates in Philosophy of Science*. Malden: Blackwell.
- Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Chalmers, David, & Jackson, Frank. 2001. Conceptual Analysis and Reductive Explanation. *Philosophical Review*, **110**(3), 315–361.
- Chalmers, David J. 2002. Does conceivability entail possibility? *Pages 145–200 of: Gendler, Tamar Szabó, & Hawthorne, John (eds), Conceivability and Possibility*. Oxford: Clarendon Press.
- Goff, Philip. 2017. *Consciousness and Fundamental Reality*. New York: Oxford University Press.
- Hill, Christopher. 1997. Imaginability, conceivability, possibility and the mind-body problem. *Philosophical Studies*, **87**(1), 61–85.
- Hill, Christopher, & Mclaughlin, Brian. 1999. There are fewer things in reality than are dreamt of in Chalmers’s philosophy. *Philosophy and Phenomenological Research*, **59**(2), 445–454.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press.
- Kirk, Robert. 1979. From physical explicability to full-blooded materialism. *Philosophical Quarterly*, **29**(116), 229–237.
- Kirk, Robert. 2005. *Zombies and Consciousness*. Oxford: Clarendon Press.
- Kripke, Saul. 1980. *Naming and Necessity*. Oxford: Basil Blackwell.

- Kung, Peter. 2010. Imagining as a guide to possibility. *Philosophy and Phenomenological Research*, **81**(3), 620–633.
- Levine, Joseph. 1983. Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, **64**(4), 354–361.
- Loar, Brian. 1990. Phenomenal states. *Philosophical Perspectives*, **4**, 81–108.
- Melnyk, Andrew. 2002. Papineau on the intuition of distinctness. *SWIF Philosophy of Mind*, **4**(1).
- Nagel, Thomas. 1974. What is it like to be a bat? *The Philosophical Review*, **83**(4), 435–450.
- Papineau, David. 2002. *Thinking about Consciousness*. Oxford: Clarendon Press.
- Papineau, David. 2007. Kripke’s proof is ad hominem not two-dimensional. *Philosophical Perspectives*, **21**, 475–494.
- Penrose, Roger. 1989. *The Emperor’s New Mind*. Oxford: Oxford University Press.
- Perry, John. 2001. *Knowledge, Possibility, and Consciousness*. Cambridge: MIT Press.
- Price, Huw. 1996. *Time’s Arrow and Archimedes’ Point: New Directions for the Physics of Time*. Oxford: Oxford University Press.
- Price, Huw. 2004. On the origins of the Arrow of Time: why there is still a puzzle about the low-entropy past. *Pages 219–239 of: Hitchcock, Christopher (ed), Contemporary Debates in Philosophy of Science*. Malden: Blackwell.
- Stoljar, Daniel. 2006. *Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness*. New York: Oxford University Press.
- Stoljar, Daniel. 2010. *Physicalism*. New York: Routledge.
- Sundström, Pär. 2008. Is the mystery an illusion? Papineau on the problem of consciousness. *Synthese*, **163**(2), 133–143.

Tye, Michael. 1999. Phenomenal consciousness: the explanatory gap as a cognitive illusion. *Mind*, **108**(432), 705–725.

Watkins, Michael. 1989. The knowledge argument against ‘the knowledge argument’. *Analysis*, **49**(3), 158–160.

Yablo, Stephen. 2008. *Thoughts: Papers on Mind, Meaning, and Modality*. Oxford: Oxford University Press.