ORIGINAL RESEARCH

CrossMark

Consciousness, Conceivability, and Intrinsic Reduction

Jonathon VandenHombergh[1]

Abstract

Conceivability arguments (Descartes in Discourse on method; and, meditations on first philosophy, Hackett Pub. Co, Indianapolis, 1993; Kripke in: Munitz (ed) Identity and Individuation, New York University Press, New York, pp 135–164, 1971; Kripke in Naming and necessity, Harvard University Press, Cambridge 1980; Jackson in Philos Stud 42(2):209–225, 1982; Chalmers in The conscious mind: in search of a fundamental theory, Oxford University Press, Oxford, 1996; Chalmers in: Chalmers (ed), The character of consciousness, Oxford University Press, Oxford, 2010) constitute a serious threat against reductive physicalism. Recently, a number of authors (Bayne in Philosophia 18:265–270, 1988; Marton in Southwest Philos Rev 14(1):131–138, 1998; Sturgeon in Matters of mind: consciousness, reason, and nature, Routledge, Abingdon, 2000; Frankish in Philos Q 57(229):650–666, 2007; Brown in J Conscious Stud 17(3–4):47–69, 2010; Campbell et al. in Philos Q 67(267):223–240, 2017; VandenHombergh in Analysis 77(1):116–125, 2017) have proven and characterized a devastating logical truth, (IN), centered on these arguments: namely, that their soundness entails the *inconceivability* of reductive physicalism. In this paper, I demonstrate that (IN) is *only* a logical truth when reductive physicalism is interpreted in its stronger, *intrinsic* sense (e.g., as an identity theory), as opposed to its weaker—yet considerably more popular—*extrinsic* sense (e.g., as a supervenience theory). The basic idea generalizes: perhaps surprisingly, *stronger* (intrinsic) forms of reduction are *uniquely* resistant to the conceivability arguments opposing them. So far as the modal epistemology of reduction is concerned, therefore, it pays to go intrinsic.

✉ Jonathon VandenHombergh
vandenhomber@wisc.edu

1 Department of Philosophy, University of Wisconsin-Madison, 5185 Helen C. White Hall, 600 North Park Street, Madison, WI 53706, USA

1 Introduction

Physicalism is the view that *phenomenal properties* (Block 1995)—those it is "like something" to consciously experience (Nagel 1974)—are reducible to physical properties or the functional properties they realize (Jackson 1998).[1] If physicalism were true, and if physical properties are explanatorily more tractable than phenomenal ones, then the myriad qualities of conscious life would be far less mysterious than they appear. The searing quality of a paper cut, for instance, would *not* be a ghostly epiphenomenon, but rather something deeply connected to the nervous activity with which it correlates.

Crucial to the physicalist's position, of course, is a disambiguation of the term "reducible." While it has well-known theoretical (Nagel 1961; Fodor 1974) and even a priori senses (Chalmers and Jackson 2001), it will be simplest here to focus on *metaphysical* reducibility. And this, broadly put, admits of both an *extrinsic* and an *intrinsic* flavor. According to the former, reducibility is equivalent to the claim that:

(Extrinsic Physicalism)	Necessarily, any phenomenal property instance is entailed by some physical property instance.

As a well-trodden (if scientifically questionable) example, an instance of *C-fiber firing* entails an instance of *searing pain* in every metaphysically possible world. According to the latter, reducibility comes to the following:

(Intrinsic Physicalism)	Any phenomenal property is identical to (is the selfsame property as) some physical property

Once again with the caricature: *C-fiber firing* is the selfsame property *as searing pain*, and similarly for other relevant property pairs. Intrinsic and extrinsic physicalism are inequivalent reductive positions on the nature of phenomenal properties. Even so, the former entails the latter, under the standard assumption that, if two individuals are the selfsame thing, they *cannot* fail to be so, and if they are *not* the selfsame thing, they *must* be so distinct (for reasons familiar from Kripke 1971, 1980). For if this assumption holds, then any particular identity witnessing intrinsic physicalism is a necessary one. And if any such identity is necessary, then, given Leibniz's Law, there can be no possible world where one instantiates and the other does not. Therefore, intrinsic physicalism entails extrinsic physicalism and, at least in some logical sense, the former is the stronger reductive doctrine.

Compelling arguments have been offered against both forms of physicalism. Most resilient among them is the class of *conceivability arguments* (Descartes 1993; Kripke 1971, 1980; Chalmers 1996, 2010; etc.), according to which the conceivable dissociation of the physical and phenomenal entails the falsity of physicalism. Abstracting from the details of each particular argument (especially, though harmlessly, from the two-dimensional machinery of recent versions; e.g., Chalmers 1996, 2010), consider the following generalized form:

---

[1] In what follows, "physical" will cover both physical and functional properties. For a range of examples from the phenomenal domain, see especially Chalmers (1996).

(Conceivability Argument)

(1)　It is conceivable that some phenomenal property fails to instantiate conjointly with any physical property.
(2)　Any proposition is conceivable only if it is possible.
(3)　Any physical and phenomenal properties are (non-)identical only if they are necessarily (non-)identical.
(4)　Therefore, extrinsic and intrinsic physicalism are false.

The conceivability argument is valid, and it threatens physicalism by its opposition to both the extrinsic and intrinsic flavors. Premise (1) asks us to accept the conceivability of a situation physically identical to, and yet phenomenally distinct from, our own—a world with C-fiber firing in the absence of pain (or in the presence of, say, a tickle). We will refer to this conceived state of affairs as a "zombie situation." It is obvious from premises (1) and (2) that, possibly, there is a zombie situation. But then it will be equally obvious that the physical does not necessitate the phenomenal, so that extrinsic physicalism is false. Of course, premise (3) is merely the Kripkean assumption about the modal status of identity. If it is true, then, as mentioned previously, intrinsic entails extrinsic physicalism. So, if the latter is false, the former is too, and the conclusion swiftly follows.

A number of recent papers (Bayne 1988; Marton 1998; Sturgeon 2000; Frankish 2007; Brown 2010; Campbell et al. 2017; VandenHombergh 2017) have pointed out that conceivability arguments *against* physicalism are apparently sound *only if* physicalism is *inconceivable*. For ease of reference, and a bit more precisely:

(IN)　If the conceivability argument is sound, then some phenomenal property isn't even *conceivably* reducible to *any* physical property.

(IN) is purportedly a *logical truth* (i.e., a tautology), the defense of which I will leave for below. It has been taken to support charges of *burden-shifting* (e.g., Marton 1998) *question-begging* (e.g., Brown 2010; VandenHombergh 2017), and *redundancy* (e.g., Frankish 2007; Campbell et al. 2017) against the proponent of the conceivability argument (against, e.g., dualists). As such, it constitutes a compelling objection to that argument, and sidesteps the need for controversial disputation over its soundness.

While I find these charges reasonable, and will largely presuppose their reasonableness in what follows,[2] it has largely escaped notice that (IN) is interpretable in two ways—one for each flavor of reduction mentioned above:

---

[2] It is worth mentioning that some philosophers have apparently bitten the bullet, arguing that physicalism *is* inconceivable (see, e.g., Chalmers 2010). The bite is then softened by an insistence that physicalism *may* be conceivable *in some sense*, albeit *not* in the sense relevant to conceivability-possibility entailment principles relied on in the defense of (IN). While the discussion here is fairly involved, it is difficult to see how this move might avoid controversy. Campbell et al. (2017) persuasively argue that it requires an argument, the existence of which undermines the purpose of any *independent* conceivability argument. VandenHombergh (2017) mentions that the "*prima facie* negative conceivability" of physicalism—offered as an olive branch in Chalmers 2010; cf. his 2002—would merely collapse to possibility-entailing senses of conceivability. For that matter, it is hard to see how a dualist

(INEP)   If the conceivability argument is sound, then extrinsic physicalism is inconceivable

(INIP)   If the conceivability argument is sound, then intrinsic physicalism is inconceivable

And, as a matter of fact, (INEP) is *not* a logical truth; only (INIP) enjoys this privilege. As such, if (IN) is taken to be a decisive objection to the conceivability argument—as, I think, it should be—then physicalists ought to endorse (INIP) over (INEP). Accordingly, physicalists should endorse intrinsic physicalism, and, by consequence, the extrinsic physicalism it entails, rather than the latter doctrine *simpliciter*.

In Sect. 2, I offer a more specific demonstration of (INIP), and I provide a counter-model against the purported validity of (INEP). Section 3 considers two attempted defenses of (INEP). These are drawn largely from the authors mentioned above: Marton (1998) and Sturgeon (2000) suggest endorsing the modal logic S5, while Frankish (2007), Brown (2010), and Campbell et al. (2017) rely on an "analytic"[3] assumption about the conditional truth of physicalism. Each of these attempts succeeds in justifying something quite like (INEP), but only at the cost of introducing new premises or disagreeable complexity. Section 4 considers two objections to the claim that intrinsic physicalism is the doctrine uniquely resistant to conceivability arguments. Finally, Sect. 5 considers the relevance of this conclusion both for the philosophy of mind and for modal epistemology in general: counterintuitively, the *stronger* the logico-modal structure of a reduction relation, the more tenuous our intuitions *against* it, and the less reliable conceivability will be in settling any number of important philosophical disputes.

2 Reductive Inconceivabilia

2.1 (INIP)

The proof of (IN) can be usefully presented by reference to a "symmetric" or "mirror argument" (Campbell et al. 2017); that is, an argument *in favor of* physicalism whose premises are as justified as those of the conceivability argument. Of course, since (IN) can be interpreted in two ways—i.e., as either (INIP) or (INEP)—we will expect their proofs to rely on two distinct sorts of mirror argument. In this section, we will consider the following mirror, which aids in the successful proof of (INIP):

(Mirror Argument)

Footnote 2 continued

might advance *both* the *prima facie* conceivability of physicalism *and* the (required) *ideal* conceivability of zombies, without reducing this central issue of consciousness metaphysics to mere intuition-mongering. Perhaps it *is* mere intuition-mongering, but less drastic measures ought first to be considered.

[3] This terminology is unique to Campbell et al. (2017), but the assumption is more or less the same (as I will discuss below).

(1)   Intrinsic physicalism is conceivable.
(2)   Any proposition is conceivable only if it is possible.
(3)   Any physical and phenomenal properties are (non-)identical only if they are necessarily (non-)identical.
(4)   Therefore, intrinsic and extrinsic physicalism are true.

Once again, the mirror argument is valid. Premises (1) and (2) jointly entail the possible identity of any phenomenal property with some physical property; i.e., the possible truth of intrinsic physicalism. But the Kripkean assumption about the modal status of identity tells us that if any two properties are not identical, then they are *necessarily* not identical. Equivalently, if they are *not* necessarily not identical— if they are *possibly* identical—then they are identical, *simpliciter*. Therefore, any phenomenal property is identical to some physical property. Since this is just intrinsic physicalism, and intrinsic entails extrinsic physicalism, it likewise follows that both reductive doctrines are true. Hence the conclusion, (4), which obviously contradicts that of the original conceivability argument.

As intended, the mirror argument is symmetric with the conceivability argument, and conversely: all of their premises, with the exception of the first, are identical. We can now prove (INIP) by way of this symmetry:

(INIP)   If the conceivability argument is sound, then intrinsic physicalism is inconceivable.

Let us suppose that the conceivability argument is sound. It follows immediately that the mirror argument is unsound, since its conclusion contradicts that of the conceivability argument. Thus, one of the mirror premises is false. Owing to the arguments' symmetry, however, it cannot be that mirror premises (2) or (3) are false, since these are the same as premises (2) and (3) of the conceivability argument, and that argument is (by hypothesis) sound. Therefore, it must be that mirror premise (1) is false. Discharging our initial assumption, it follows that if the conceivability argument is sound, then intrinsic physicalism is inconceivable; (INIP) is tautologically true.

It will be useful here to clear up two possible confusions. The first is a confusion by parity. It would seem that, given largely similar reasoning, the soundness of the mirror argument would entail the inconceivability of the zombie situation. To the extent that the zombie situation and dualism are materially equivalent, we might be able to prove the mirror of (INIP):

(PINI)   If the mirror argument is sound, then dualism is inconceivable.

And, indeed, we *can* prove (PINI). For suppose that the mirror argument is sound. It will follow that the conceivability argument is unsound; symmetry guarantees that the source of unsoundness must be the latter's first premise, thus entailing the inconceivability of the zombie situation. And this is just to say that we cannot conceive of the zombie situation, or the truth of dualism, if the mirror argument is sound; (PINI) is tautologically true. The confusion is then easy to state: to the extent that (INIP) is *problematic* for the dualist, and so appears to impugn her position,

(PINI) is problematic *for the physicalist*, and so impugns hers as well. We therefore may have proved too much. But this is not the case, for two reasons.

First, and more weakly, we will recall that it was the *dualist* who offered the conceivability argument in opposition to physicalism, rather than the physicalist offering the mirror argument in opposition to dualism. Indeed, the physicalist has not even *suggested* that the mirror may provide a sound counterargument to dualism. Thus, while (PINI) is indeed a *logical* truth, it is not one that problematizes physicalism, because the physicalist has not endorsed the truth of its antecedent. On the contrary, the dualist *has* asserted the truth of (INIP)'s antecedent, and is therefore committed to the inconceivability of intrinsic physicalism. This latter commitment is problematic, for the reasons mentioned above.

Second, and more strongly, the dualist apparently *must* offer the conceivability argument in opposition to physicalism, whereas the physicalist *need not* offer the mirror argument against dualism. After all, the physicalist has a variety of arguments for her position which are obviously distinct from the mirror argument: e.g., arguments from mental causation and physical closure (e.g., Kim 1989, 2011, ch. 6; Bennett 2007), phenomenal judgment (Rudd 2000; Kirk 2005, ch. 4; Neagle 2012),[4] parsimony (Smart 1959; Brandt and Kim 1967; Sober 2015, ch. 5), induction on the successful history of reductive science (Smart 1959), and so on. The dualist, however, appears only to have canonical arguments nearly equivalent to the conceivability argument: e.g., Jackson's "knowledge argument" (1982), Block's "Chinese nation" (1972), certain interpretations of Searle's "Chinese room" (1980), and of course the alternative conceivability arguments of Plantinga (1974, ch. 4); Gertler (2008), and so on. And it is clear that the core intuition underlying *each* of these is nearly indistinct from that of the conceivability argument; i.e., the intuition that phenomenal consciousness is conceivably dissociated from the physico-functional world. By "nearly equivalent" and "nearly distinct," I mean only that these arguments are just as susceptible to objections like (INIP), even if, strictly speaking, they are distinct arguments (in this vein, see Brown 2010; Campbell et al. 2017, each of whom considers analogous objections to the knowledge argument). As such, not only is (INIP) problematic for the dualist who *actually* endorses the conceivability argument, it is apparently problematic for the dualist who *doesn't* endorse it but endorses some *other* anti-physicalist argument. Neither is true of the physicalist, however. Therefore, (PINI)—while a logical truth—is entirely unproblematic.

The second possible confusion is simpler to clear up. One may reasonably point out that the conceivability and mirror arguments are *not* symmetric, on the following grounds: the Kripkean third premise, concerning the (non-)necessity of (non-)identity, is tacitly a conjunction of *two* premises. The *first* says that any true identity is a necessarily true one. The *second* says that any false identity is a necessarily false one. Furthermore, the conceivability argument uses only the first— and the mirror argument, only the second—in order to prove their corresponding

---

[4] Rudd and Naegle do not explicitly conclude in favor of reductive physicalism, given various background assumptions. Nevertheless, both arguments are opposed to epiphenomenalism, and, given the abandonment of certain assumptions, could work in favor of the reductive position (cf. Chalmers 1996, ch. 5).

(and inconsistent) conclusions. We might then wonder: couldn't the dualist deny the second tacit premise, thereby avoiding the soundness of the mirror argument without needing to claim that *anything* is inconceivable?

Logically, the answer to this question is "yes." The dualist could break symmetry by denying the necessary falsity of any non-identity (see Bayne 1988 for a similar discussion). But there is here a symmetry of *justification* which makes this move entirely inadmissible. The standard Kripkean justification (ignoring, e.g., the technical concept of rigid designation) says that identity is, at root, a relation of *selfsameness*, and that selfsameness comes beholden with second-order modal properties. If *x* and *y* are identical—if they just *are* the *selfsame thing*—then it would be absurd to accept the counterfactual possibility of their (its!) non-selfsameness. But there is nothing about the *actual world* which seems uniquely to support this reasoning. For even if *x* and *y* merely *could have been* the selfsame thing, they would be the selfsame thing in *some world*, and it would be equally absurd to think that what is *somewhere* the selfsame thing could *fail* to be itself here. Modal space is in this regard egalitarian. But the second of these inferences—from possible selfsameness to selfsameness—is merely the contraposition of the second tacit premise. Unless the actual world is unique with respect to its selfsame members, there is no reason for abandoning the first tacit premise which wouldn't require abandoning the second, and conversely. Rhetorically speaking, the symmetry is maintained.[5]

Having cleared up these two confusions, let us now proceed to the discontents of (INIP)'s extrinsic alternative.

2.2 (INEP)

Recall (INEP), the claim that

(INEP) If the conceivability argument is sound, then extrinsic physicalism is inconceivable.

If (INEP) were a logical truth, then the conceivability argument would appear deficient against both intrinsic *and* extrinsic physicalism, and thus against physicalist reduction generally. And, indeed, a number of authors have argued in favor of (INEP) or given arguments justifying its near-equivalents (e.g., Marton 1998; Sturgeon 2000; Frankish 2007; Brown 2010; Campbell et al. 2017). While (INEP) may indeed be *true*, however, it is *not* obviously a *logical* truth, and so tacitly requires a premise to which the dualist might object. Equivalently, the demonstration of (INEP) requires that we *break* the symmetry of the mirror argument, which allows the dualist room to deny the charges of burden-shifting, question-begging, and redundancy mentioned above. We will consider possible tacit premises in the next section. For the moment, I would like to close Sect. 2 by

---

[5] This symmetry is reflected in the details of the corresponding formalism: to capture the semantic *rigidity* of terms in a first-order modal language, one must endorse the conjunctive axiom schema $((s = t \supset \Box(s = t)) \land (s \neq t \supset \Box(s \neq t)))$ (cf. Garson 2013).

producing a counter-model for (INEP)—a model according to which the conceivability argument is sound yet extrinsic physicalism *is* conceivable—thereby refuting its tautological pretense.

A counter-model for (INEP) may be given as follows. Let $\{w_@, w_1, w_2\}$ be a set of *possible worlds*, and let $\{z, e\}$ be a set of *conceivable worlds*. The domain of any world from any set is identical—{*C-fiber firing*, *searing pain*}—letting the physical and phenomenal quantifiers in the conceivability argument range over their respective properties (e.g., "any phenomenal property" ranges over *searing pain* but not *C-fiber firing*). In spite of this, worlds will differ over which properties they *instantiate*. For simplicity, then, we will identify worlds with the set of their instantiated properties:

$w_@ =$ {*C-fiber firing*, *searing pain*}
$w_1 =$ {*C-fiber firing*}
$w_2 =$ {*C-fiber firing*, *searing pain*}
$z =$ {*C-fiber firing*}
$e =$ {}

Generally speaking, a proposition of the form "it is conceivable that φ" is true in any world x (of any kind) if and only if it is true in some conceivable world c and c is accessible from x; *mutatis mutandis* for possibility and necessity. To complete the model above, therefore, we require the following facts about accessibility: the actual world, $w_@$, accesses z, e, and w_1; w_1 accesses w_2; and e accesses w_2. This model makes every premise of the conceivability argument *and* the conceivability of extrinsic reduction true in the actual world, $w_@$. As such, it demonstrates that (INEP) is not a logical truth. The proof is as follows.

Immediately, we see that both *searing pain* and *C-fiber firing* are non-identical in every world of the model. Since there are necessarily no true identities at all, the conceivability argument's third premise is vacuously true in $w_@$.

We see next that z is a conceivable world in which only *C-fiber firing* is instantiated (so that, trivially, *searing pain* is not). As these are the only existing physical and phenomenal properties, the following will be true in z: some phenomenal property fails to instantiate conjointly with any physical property (e.g., the zombie situation is true). Since $w_@$ accesses z, it will be true in $w_@$ that, conceivably, the zombie situation obtains. And this is just to say that the premise (1) of the conceivability argument is true there.

Next, note that e accesses only world w_2. As such, any proposition true in w_2 will be necessarily true in e. And since it is true in w_2 that *C-fiber firing* and *searing pain* are instantiated, it is necessary in e that *C-fiber firing* instantiates only if *searing pain* does. Once again, these are the *only* physical and phenomenal properties in any world. It will therefore follow that, in e, *any* phenomenal property is such that there is *some* physical property necessitating it. More simply, extrinsic physicalism will be true in the conceivable world e. And since $w_@$ accesses e, extrinsic physicalism will be conceivable in $w_@$.

Finally, to demonstrate the truth of the conceivability argument's second premise, we must show that it is satisfied for every proposition in the model. To do this, we divide the set of propositions in half. The first half contains those which are

conceivably true in $w_@$ according to the model; each of these propositions must also be *possible* in the actual world (else they will falsify the conceivability argument's second premise). And indeed they are. On the one hand, we have the proposition which witnesses the zombie situation. Of course, world w_1 makes precisely the zombie situation true, and it is accessible from the actual world. Therefore, this proposition is both conceivable and possible in the actual world. On the other hand, we have the proposition representing extrinsic physicalism. We note first that instances of *C-fiber firing* entail those of *searing pain* in w_2. Once again, because all worlds contain exactly the same properties, it will be true in w_2 that, for *any* phenomenal property, there is *some* physical property entailing it. Since w_1 accesses w_2, and w_2 is the *only* world w_1 accesses, it is true in w_1 that the prior entailment is *necessary*; that is, extrinsic physicalism is true in w_1. But the actual world accesses w_1. Therefore, extrinsic physicalism will *actually* be possible, and so it is both conceivable and possible in the actual world. It follows that premise (2) of the conceivability argument is true in the actual world, for every proposition which is likewise conceivable there. The second half of the set of propositions contains all those which are *not* conceivable in $w_@$ according to the model. Each of these propositions makes the second premise vacuously true in the actual world (Fig. 1).
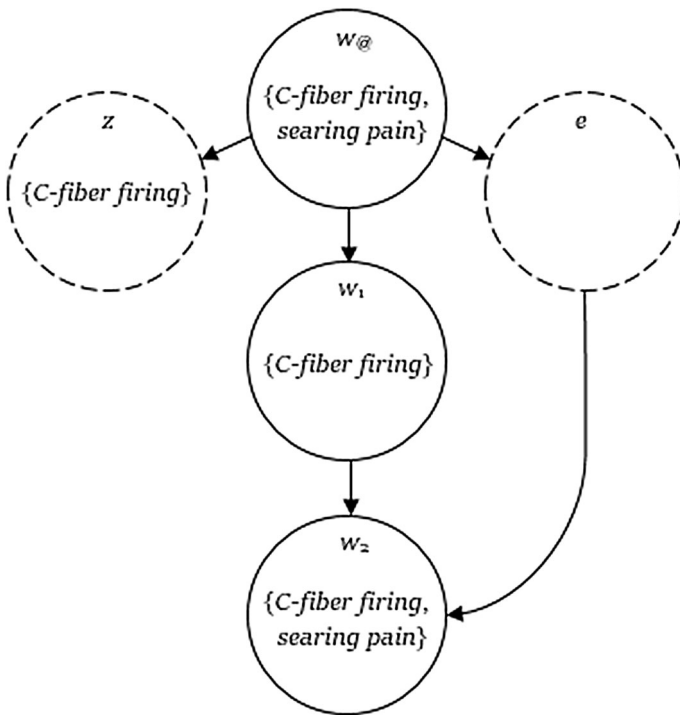

Fig. 1 A counter-model to (INEP). Possible worlds in *W* are solid circles, conceivable worlds in *C* are dashed circles, arrows represent accessibility, and each property set in each world is instantiated in that world

Therefore, that premise is true in the actual world, and our proof is complete.[6]

3 Whither Extrinsic Physicalism?

Despite the demonstrations above, a number of authors have defended propositions akin or equivalent to (INEP). This is because, as I will now argue, (INEP) *can* be made true if we are willing to *introduce* a tacit premise ψ into the mirror argument (as each of these authors do). Equivalently, while (INEP) is not a logical truth, it can be *augmented* into one thus:

(INEP$_\psi$)   If the conceivability argument is sound *and* some ψ is true, then extrinsic physicalism is inconceivable.

I will here consider two pre-existing interpretations of ψ: the first, *modal* (in Marton 1998; Sturgeon 2000), and the second, *analytic* (in Frankish 2007; Brown 2010; Campbell et al. 2017). Each interpretation *does* satisfy (INEP$_\psi$); nevertheless, I will wish to argue against both, and therefore against the use of (INEP$_\psi$) as a general strategy for opposing the conceivability argument. For, in addition to some independently problematic consequences of either interpretation, the introduction of ψ destroys the symmetry which made (IN) so compelling. The dualist is therefore free to deny ψ, so that (INEP$_\psi$), unlike (INIP), fails to support the charges of burden-shifting, circularity, and redundancy.

3.1 Modal Interpretations of ψ

Both Peter Marton and Scott Sturgeon have remarked, in more or less equivalent terms, that (INEP) is a logical truth: if the conceivability argument is sound, then extrinsic physicalism is inconceivable. As a matter of fact, they have only demonstrated the validity of (INEP$_\psi$) under its modal interpretation:

(INEP$_5$)   If the conceivability argument is sound and the axiom (5) is true, then extrinsic physicalism is inconceivable,

where ψ has been identified with the axiom

(5)   Any proposition is possibly necessary only if it is necessary

---

[6] Note: we might have equally let the actual world instantiate no properties at all, and the model still would have succeeded. Nevertheless, I included the (slightly more complex) model in order to show that the invalidity of (INEP) is consistent with what all (non-eliminative and non-idealist) parties to the dispute accept: there are at least *some* physical and phenomenal properties which *actually* instantiate.

Consider Marton's argument first (supplanting his definition of physicalism, "MT,"[7] with extrinsic physicalism):

"(1)   If zombies are possible then [extrinsic physicalism] is false.
 (2)   Therefore, [extrinsic physicalism] is false.
          …
 (3)   If [extrinsic physicalism] is false then [extrinsic physicalism] is necessarily false.
 (4)   Therefore, [extrinsic physicalism] is necessarily false.
          …
 (5)   If [extrinsic physicalism] is necessarily false, then it is impossible.
          …
 (6)   If [extrinsic physicalism] is impossible then it is inconceivable" (1998).

It is easy to see that if zombies are possible, then extrinsic physicalism is impossible (given premises (1), (3), and (5), above). By Marton's sixth premise, it will likewise follow that if zombies are possible, then extrinsic physicalism is inconceivable. And since it is precisely the conceivability argument which establishes the possibility of zombies, Marton seems to have shown that the conceivability argument is sound only if extrinsic physicalism is inconceivable. In other words, he has appeared to have proven (INEP). Of course, he has only done so on the assumption that his third premise is *also* true. And this premise, as a matter of fact, is just an instance of the axiom (5) mentioned above. After all, premise (3) says that if extrinsic physicalism is false, then it is necessarily false. This is of course equivalent to the claim that if extrinsic physicalism is not necessarily false, it is true, which itself amounts to the claim that if extrinsic physicalism is *possible*, it is true. But extrinsic physicalism is *itself* a claim about necessary entailment. As such, Marton's premise (3) says that if it is *possibly necessary* that any phenomenal instance is entailed by some physical one, then it is *necessarily* thus entailed. And this is clearly an instance of (5).

Sturgeon reasons similarly, and the point doesn't need belaboring. We note only that he considers the conceivable *impossibility* of a zombie situation ("Z"), from which he infers (with some notational changes) that:

---

[7] In his words, MT is defined as the thesis that "All facts of (phenomenal) consciousness logically supervene on the totality of fundamental (micro-) physical facts" (Marton 1998). MT is equivalent to extrinsic physicalism, so long as: "facts" are understood as "property instances," "logical supervenience" is understood as "metaphysical supervenience" (itself understood as metaphysical necessitation), and "the totality of… physical instances" is understood as the instantiation of some complex physical property.

"(1) [It is possibly impossible that] Z.
Yet the logic now entails
(2) [It is impossible that] Z" (2000).

In other words, because it is *conceivable* that the zombie situation is impossible, it is *possibly* impossible—premise (1) above. After all, conceivability entails possibility, according to the dualist. Yet Sturgeon's inference, from premise (1) to (2), is once again underwritten by an instance of axiom (5). After all, impossibility is necessary falsity, so that (1) entails (2) if and only if the possibly necessary falsehood of Z entails the necessary falsehood of Z; i.e., if and only if the possible necessity of not-Z entails the necessity of not-Z. But this would be guaranteed if, for any proposition (negated or otherwise), possible necessity entailed necessity *simpliciter*—a guarantee honored by S5.[8] It is easy enough to see, furthermore, that the impossibility of the zombie situation is equivalent to the falsity of extrinsic physicalism.

This suggests that Sturgeon has relied on exactly the same mirror argument as Marton, which is to say, he has demonstrated that (INEP₅) is a logical truth.

Abstracting from the particularities of either argument, then, Marton and Sturgeon both appear to deploy the following mirror:

(Modal Mirror Argument)

(1)   Extrinsic physicalism is conceivable.
(2)   Any proposition is conceivable only if it is possible.
(3)   Any proposition is possibly necessary only if it is necessary.
(4)   Therefore, extrinsic physicalism is true.

The modal mirror is expectedly valid. By the usual symmetric reasoning, it will follow that if the conceivability argument is sound, then *either* (5) is false, or modal mirror premise (1) is false—that is, either (5) is false or extrinsic physicalism is inconceivable. Equivalently, if the conceivability argument is sound *and* (5) is true, then extrinsic physicalism is inconceivable. And this is exactly the modal interpretation of (INEP$_\psi$), promised above.

While Sturgeon leaves the (5) assumption tacit, Marton explicitly addresses its use [under the heading S5—the modal system characterized by (5)]. However, he offers that the dualist

needs S5 for the success of his argument. Imagine that axiom-schema S5 does not hold. According to the Kripke-style semantics, there might be then some possible world which is not accessible from the actual one. It is not impossible then that materialism is true in the actual world and all the worlds accessible from it, nonetheless, some of the worlds, not available from the actual world, are in fact zombie worlds. (1998).

---

[8]   VandenHombergh (2017) incorrectly describes Sturgeon's assumption as an instance of S4, or the transitive 4 axiom. Sturgeon *earlier* appeals to S4, in order to demonstrate that the possible possibility of a zombie situation entails its possibility, *simpliciter*—but this concerns his own reconstruction of the conceivability argument, not its mirror.

If Marton is correct, then the symmetry would indeed be reestablished, and the dualist could not simply opt against the axiom (5). But, while I of course share his suspicions against the conceivability argument, it is clear enough that the dualist *does not* require (5) (and therefore does not require S5). At no point in the proof of Sect. 1 was (5) appealed to, and there is (I think) a diagnosable error in the quotation above. For *if* it is *actually* true that a zombie situation is conceivable, and *if* it is *actually* true that conceivability entails possibility, then it is *actually* possible that a zombie situation exists. Under the Kripke-style semantics to which Marton appeals, this *requires* there to be an *accessible* zombie world (whatever inaccessible worlds there might otherwise be). As such, the truth of the conceivability argument's first two premises block Marton's claim about accessibility, and they do so without need of S5.[9] This is perhaps why Chalmers, in his defense of (one form of the) conceivability argument, floats the following easy reply: "one could respond by denying S5." Indeed one could, and the symmetry promised by (IN) is therefore broken in (INEP$_5$).[10]

Marton and Sturgeon draw slightly distinct conclusions from the validity of the modal (INEP$_5$). The former, for instance, concludes variously that the burden of proof is now on the dualist to explain *why* extrinsic physicalism must be inconceivable—a consequence of the conceivability argument's soundness. He also offers that the entailment from conceivability to possibility (captured in the conceivability argument's second premise) is false, insofar as extrinsic physicalism *and* the zombie situation are both conceivable. Sturgeon draws a slightly weaker conclusion, saying that

> Full conceptual reflection sides neither with [the possibility of zombies] nor with its negation. It symmetrically endorses each side. It thereby undercuts each side (Sturgeon 2000).

But it should now be clear that the dualist needn't incur any serious burden, nor that she must abandon the entailment from conceivability to possibility, nor that she undercuts her own position. For it is well within her rights to simply deny the (5) axiom, without obvious contradiction or concern, all the while continuing to promote the conceivability argument as a sound demonstration of dualism. As such, she is not committed to the charges of burden-shifting, question-begging, or redundancy mentioned above.

---

[9] Indeed, on a technical note, it doesn't seem to me that the denial of (5) necessitates the existence of a world inaccessible from the actual one. Again, let $w_@$ access w_1, w_2, w_3, and itself, and let these be the only existing worlds. If w_1 accesses w_2 and w_3, while the latter two fail to access one another, then the accessibility relation will be non-Euclidean and the (5) axiom will be invalid—even though, by hypothesis, our world accesses everything.

[10] There are other problems. The conceivability of extrinsic physicalism, as used in the modal mirror argument, requires embedding modal operators in the scope of the conceivability operator, complicating things significantly and contradicting an otherwise plausible hypothesis in modal epistemology (cf. Chalmers's "modal rationalism" 2002).

3.2 Analytic Interpretations of ψ

A number of authors (most explicitly, Campbell et al. 2017; though cf. Frankish 2007 and Brown 2010) have taken notice of the problems above. In part to avoid them, they have chosen to defend an analytic interpretation of (INEP$_\psi$):

(INEP$_A$)   If the conceivability argument is sound and (A) is true, then extrinsic physicalism is inconceivable,

where (A) is a purported analytic truth. To articulate (A), let us first define a *physical situation* by the following proposition:

(PS)   Any phenomenal property is conjointly instantiated with some physical property.

In other words, a physical situation is a non-modal, physical state of affairs *containing* any arbitrarily specified phenomenal property. The symmetric arguments to be considered, and the analytic premise (A), both implicate physical situations. For instance, Frankish entertains the conceivability of "anti-zombies": "beings which are bare physical duplicates of us, inhabiting a universe which is a bare physical duplicate of ours, but none the less having exactly the same conscious experiences as we do" (2007). Relatedly, Brown considers the conceivability of a "shombie:" "a creature that is micro-physically identical to me, has conscious experience, and is completely physical… in a world that is stipulated to lack nonphysical properties" (2010). And, most recently, Campbell et al. 2017 discuss the conceivability of a "non-modal claim… $PT \wedge Q$," where PT is just the instantiation of (only) all physical properties and Q is the instantiation of an arbitrary phenomenal property (following the account given in Chalmers 2010). The analytic premise is then given as

(A)   (PS) is possible only if extrinsic physicalism is true.

That is, the physical situation is possible only on the condition that extrinsic reduction is true. Before discussing the merits of (A)—and some details about the *totality* of (PS), which is presently simplified—let us first recover the symmetric argument underlying (INEP$_A$).

Each author explicitly offers their own symmetric argument, from which we can abstract to a general mirror as before. Again with slight notational changes (for clarification):

"(1)   Anti-zombies are conceivable.
  (2)   If anti-zombies are conceivable, then anti-zombies are possible.
  (3)   If anti-zombies are possible, then consciousness is physical.
  (4)   So consciousness is physical" (Frankish 2007).

"(1)   P and Q [a shombie situation] is conceivable.
  (2)   If (P & Q) is conceivable, then (P & Q) is possible.
  (3)   If (P & Q) is possible then dualism is false.
  (4)   Therefore dualism is false" (Brown 2010).

"(1) [Conceivably]$(PT \land Q)$.
(2) [Conceivably]$(PT \land Q) \rightarrow$ [possibly]$(PT \land Q)$.
(3) [Possibly]$(PT \land Q) \rightarrow$ [necessarily]$(PT \rightarrow Q)$.
(4) [Necessarily]$(PT \rightarrow Q)$" (Campbell et al. 2017).

It is clear that each first premise embodies the conceivability of the physical situation. Similarly, each second premise is merely an instance of the conceivability argument's second premise, endorsing an entailment from conceivability to possibility. Thirdly, each arguments' third premise is equivalent to the analytic premise (A). This is obvious in the case of the third premise (3) (again reading PT and Q as above). If we understand "physicalism" as extrinsic physicalism, and dualism as its negation, then the first and second premise (3)'s are likewise equivalent to (A).

The authors' symmetric arguments suggest that $(INEP_A)$ relies on the following general mirror:

(Analytic Mirror Argument)

(1) (PS) is conceivable.
(2) Any proposition is conceivable only if it is possible.
(A) (PS) is possible only if extrinsic physicalism is true.
(4) Therefore, extrinsic physicalism is true.

Need I mention that this mirror argument is valid? Perhaps it is still worth working through its deployment in the proof of $(INEP_A)$, if only because *that* argument requires one further premise. As usual, we can see that the conceivability argument is sound if and only if the analytic mirror isn't, so that, if the lefthand side obtains, the mirror's first premise is false. Yet the falsity of this premise does *not* amount to the inconceivability of extrinsic physicalism, but rather the inconceivability of the (weaker) physical situation (PS). To demonstrate the logical truth of $(INEP_A)$, therefore, we are in need of one further premise:

(A*) Extrinsic physicalism is conceivable only if (PS) is conceivable.

As a matter of fact, the aforementioned authors do not rely on anything like $(A*)$; they appear to find the inconceivability of (PS) problematic enough in its own right.[11] But I will have nothing critical to say about $(A*)$ anyway, so let us grant it without argument. Having done so, it is clear that the soundness of the conceivability argument entails the inconceivability of extrinsic physicalism. Therefore, $(INEP_A)$ is logically true.

It is fairly clear that $(INEP_A)$ succeeds where its modal precursor fails. After all, the physical situation is explicitly *non-modal*; it follows that the analytic mirror premises (1) and (2) entail only that the physical situation is *possible*. This is in stark contrast to the premises of the modal mirror argument, which entailed the possibility

---

[11] With the exception of Campbell et al. (2017), who introduce somewhat distinct background premises to similar effect.

of extrinsic physicalism, and so the possible *necessity* of some proposition—it was this modal embedding which prompted the addition of the (5) axiom and gave the dualist an easy exit. But might the dualist *similarly* abandon the new premise (*A*)? The answer is complicated. Under a specific interpretation of the physical property instantiated in the physical situation, (*A*) appears to be a strictly undeniable truth. As such, the dualist *would* be committed to (*A*) *given* this interpretation. Nevertheless, as we will now take time to consider, the relevant interpretation is needlessly strong; weakening it sufficiently, however, makes (INEP$_A$) *equivalent to* the intrinsic case (INIP).

The "relevant interpretation" of the physical situation (PS) involves the concept of a "total physical duplicate." Let us say that $p_@$ is the conjunction of every physical property *individually* instantiated in the actual world. In other words, if *C-fiber firing*, *action potentiation*, *mental representation*, etc. are *each* instantiated in the actual world, then each of these properties are a (conjunctive) part of $p_@$. Then, let $t(p_@)$ be the *totalization* of $p_@$: $t(p_@)$ is instantiated in a world w if and only if $p_@$ is instantiated in w and, for any other property p^*, p^* is instantiated in w only if p^* is a part of $p_@$. Any world instantiating $p_@$ is a *physical duplicate* of the actual world. It contains all those physical properties which actually instantiate, *and possibly more*. And any world instantiating $t(p_@)$ is a "total" (Jackson 1998; Chalmers 2010) or "bare physical duplicate" (Frankish 2007) of the actual world. It contains all those physical properties which actually instantiate, *and that's all*. Let us assume, somewhat controversially but with no harm to the present point, that $t(p_@)$ still counts as a physical property. Having done so, we can understand the physical situation as quantifying over the physical property $t(p_@)$. That is, we can read (PS) as

(PS$_T$)   Any phenomenal property is conjointly instantiated with the physical
                  property $t(p_@)$.

Reflection on the quotations about "anti-zombies," "shombies," and the like, suggest that (PS$_T$) is the correct reading of the physical situation, (PS): namely, as a *total physical duplicate of our own world*, *yet containing consciousness nonetheless*.

And, indeed, it would appear that (PS$_T$) *does* make (*A*) an analytic truth—or, at any rate, something near enough. In particular, (*A*) will follow, under this totalized interpretation, *assuming the Kripkean premise about identity*. For assume the antecedent of (*A*), totalized: i.e., it is possible that (PS$_T$). If this is possible, then there is a world instantiating $t(p_@)$ and any arbitrarily chosen phenomenal property (say, *searing pain*). Equivalently, there will be a total physical duplicate w of the actual world which contains searing pain. By the definition of "totalization," however, we will have it that any property instantiated in w is a part of $p_@$. Therefore, *searing pain* will be a part of $p_@$ in w. But if *searing pain* is a part of $p_@$, then, by definition, *searing pain* is a(n actually instantiated) physical property. Trivially, therefore, it is identical in w to some physical property, p^* (where p^* just *is searing pain*). It follows that *searing pain* is *possibly* identical to p^*. And given the Kripkean assumption, *searing pain* will be *necessarily* identical to p^*. We have already seen that if two properties are Kripkean-identical, then it is necessary that one instantiates only if the other does, too. Consequently, we have it that extrinsic physicalism is true (whether or not it *too* is totally interpreted). So the possibility of

(PS_T) entails extrinsic physicalism, and (*A*) is true analytically—or, at least, true in virtue of the Kripkean properties of identity and the concept of total physical duplication. This reasoning appears to reflect the arguments given in favor of (*A*) (e.g., in Frankish 2007, p. 654; Campbell et al. 2017, p. 227).

If this is the intended justification of (*A*), it suggests that the mirror argument *really* boils down to the following:

(Totalized Analytic Mirror Argument)

(1) (PS_T) is conceivable.
(2) Any proposition is conceivable only if it is possible.
(3) Any physical and phenomenal properties are (non-)identical only if they are necessarily (non-)identical.
(4) Therefore, extrinsic physicalism is true.

Notice that we have merely given the original analytic mirror its intended totalized reading and replaced (*A*) with the original Kripkean identity premise.[12] And if this *is* the reasoning underlying the authors' arguments above, then the usual symmetric inferences suggest that ($INEP_A$) is *really* equivalent to

($INEP_A$*)   If the conceivability argument is sound, then (PS_T) is inconceivable.

And yet, as mentioned above, ($INEP_A$*)—while harder to deny—is needlessly strong.

First, recall the meaning of (PS_T): any phenomenal property is conjointly instantiated with the physical property $t(p_@)$. By the definition of totalization, this is equivalent to the claim that any phenomenal property is instantiated alongside the physical duplicate $p_@$, and any instantiated property is identical to some physical property contained in $p_@$. But note: since *any* instantiated property includes any instantiated *phenomenal* property, the latter conjunct immediately entails intrinsic physicalism![13] As such, (PS_T) entails intrinsic physicalism *and* further propositional content—namely, that the *rest* of $p_@$ is instantiated. ($INEP_A$*) therefore tells us that if the conceivability argument is sound, then *both* intrinsic physicalism *and* the instantiation of $p_@$ are inconceivable. And it is clear that the latter is doing no work beyond that done by the former—as our earlier demonstration of (INIP) makes clear, the inconceivability of intrinsic physicalism is problematic enough in its own right. Indeed, the present considerations demonstrate that extrinsic physicalism plays no role whatsoever in the proof of ($INEP_A$*). A focus on ($INEP_A$*) over (INIP) has only one effect: to needlessly introduce the conceivability of the remaining physical situation, $p_@$. Not only is this additional conceivabilium unnecessary

---

[12] There is one tacit step. The intended interpretation proves the *totalized* instance of extrinsic physicalism; existentially generalizing gets us the conclusion above.

[13] It might be objected that totalization could obtain in virtue of non-identification; e.g., by way of *extrinsic* physicalism. But this would clearly embed a necessity operator in the physical situation, therefore regenerating the problems associated with the modal interpretation. Frankish is right to suggest that totalization obtains "in virtue of… token identities"—but this, and not the resulting "metaphysical supervenience," is what does all the heavy-lifting (2007).

(given the availability of (INIP)), its complexity might allow a tenacious enough dualist to argue that $p_@$ *is* inconceivable after all—thus allowing them to embrace (INEP$_A$*) without incurring the charges of burden-shifting, question-begging, or redundancy. Of course, these problems would be entirely avoided if we simplified (INEP$_A$*) to the following: if the conceivability argument is sound, then intrinsic physicalism, *and only (intrinsic physicalism)*, is inconceivable. And this simplification *is* available. It is nothing other than (INIP), the preferred conditional of this paper.

4 Objections[14]

There are (at least) two salient objections worth addressing. First, notice that the conceivability argument is, in some sense, a conjunction of two distinct arguments—one for each form of physicalism on offer. The first argument, made explicit below, directly opposes intrinsic physicalism

    (Intrinsic Conceivability Argument)

(1) It is conceivable that some phenomenal property fails to instantiate conjointly with any physical property.
(2) Any proposition is conceivable only if it is possible.
(3) Any physical and phenomenal properties are (non-)identical only if they are necessarily (non-)identical.
(4) Therefore, intrinsic physicalism is false.

By familiar reasoning—in short, the proof of (INIP)—the intrinsic conceivability argument will entail the inconceivability of intrinsic physicalism. Were a dualist to offer this argument against intrinsic physicalism, therefore, it would be reasonable for a physicalist to reply that the argument is question-begging (or otherwise fallacious)—to reply, in other words, that (INIP) is too great a cost. But the second argument, in direct opposition to extrinsic physicalism, looms:

    (Extrinsic Conceivability Argument)

(1) It is conceivable that some phenomenal property fails to instantiate conjointly with any physical property.
(2) Any proposition is conceivable only if it is possible.
(3) Therefore, extrinsic physicalism is false.

What could the proponent of intrinsic reduction say against the extrinsic conceivability argument? Such a physicalist cannot ignore the argument, because intrinsic physicalism *entails* extrinsic physicalism, and thus the extrinsic argument refutes the former just as much as the latter. Furthermore, this argument does not

---

[14] I am grateful to an anonymous reviewer from *Erkenntnis* for posing these serious objections.

seem to be question-begging, since it makes no mention of intrinsic physicalism whatsoever—indeed, and again for familiar reasons, the closest it *seems* to come to any such error is the contentious (INEP). Consequently, while it may be true that the original conceivability argument begs the question against intrinsic physicalism, the extrinsic conceivability argument refutes that doctrine apparently without any similar cost. And if both intrinsic and extrinsic physicalism fall to non-fallacious arguments, then the main thesis of this paper is incorrect: the former is *not* preferable to the latter.

It may seem as if the natural response is to simply abandon the entailment from intrinsic to extrinsic physicalism—to embrace a radical reduction constituted by supervenience-unsupportive identities. Yet this kind of intrinsic reduction really *does* seem inconceivable! Luckily, there is an alternative. Recall that the intrinsic-extrinsic entailment depends on the Kripkean assumption about identity—effectively, premise (3) of the intrinsic conceivability argument given above. Therefore, if the *extrinsic* argument is to defeat not *just* extrinsic physicalism, but intrinsic physicalism *as well*, it must *also* grant premise (3). And in such a case, the premises of the extrinsic argument become equivalent to those of the intrinsic argument. To the extent that the latter argument is question-begging or otherwise fallacious, then, the former will be too. Of course, the dualist may simply deny the Kripkean assumption (contrary, perhaps, to the inclination of all but the most anomalous of intrinsic physicalists). In such a case, however, the dualist will have *no* argument against intrinsic physicalism—the original, extrinsic, and intrinsic conceivability arguments will *all* fail to controvert intrinsic physicalism. Consequently, there is no situation in which the extrinsic argument *both* defeats intrinsic physicalism *and* fails to beg the question (i.e., fails to validate (INIP)). The preferability of intrinsic physicalism is, therefore, sustained.

But there is a second objection. Let us imagine that an extrinsic physicalist has been presented with the extrinsic conceivability argument, absent any premise about Kripkean identities. Let us further suppose that this physicalist has become convinced that (INIP) is valid, and that it evinces some fallacy in the *intrinsic* conceivability argument. The core suggestion of this paper has been that such a physicalist should adopt intrinsic physicalism—thus side-stepping the extrinsic argument and inoculating against the intrinsic argument with (INIP). Yet there is admittedly a sense in which this move is *itself* question-begging. For now, one has responded to the extrinsic argument by adopting a position (intrinsic physicalism) *logically stronger* than the position opposed by that argument (extrinsic physicalism). And how could one *rationally* object to *any* argument by adopting a position logically stronger than the one opposed by that argument? Suppose an atheist produced an argument against the existence of God, and a theist responded by embracing the doctrine of *super-theism*: belief in a being whose existence entails the existence of God. This move would rightly be considered suspicious—it is akin to embracing the conjunction of some new position with the negation of one of the atheist's premises. How, then, is intrinsic physicalism any different from super-theism?

On the one hand, the objection is disarmed by the response to its predecessor. Intrinsic physicalism is *not* logically stronger than extrinsic physicalism *unless* the

Kripkean assumption is granted. And *if* it is granted, the extrinsic conceivability argument simply *becomes* the question-begging intrinsic conceivability argument, *whatever* the physicalist chooses to believe. On the other hand, however, there are relevant disanalogies between intrinsic physicalism and, e.g., super-theism (construed as argumentative strategies against dualism and atheism, respectively). Intrinsic physicalism has ample independent support, as discussed briefly above; super-theism has none. Part of this support is abductively related to extrinsic physicalism, as the existence of true psycho-physical identities would *explain* psycho-physical necessitation—the *reason* that the phenomenal is necessitated by the physical is *because* intrinsic physicalism is true. Indeed, it is frequently observed that extrinsic physicalism is explanatorily quite weak, much as mere statistical correlation is explanatorily weaker than causation (see, e.g., Kim's 1993 comments on supervenience). It would make good sense, then, for an extrinsic physicalist to *independently* adopt intrinsic physicalism; a mere byproduct of this, as demonstrated above, is inoculation from certain conceivability arguments. For these reasons, the adoption of intrinsic physicalism needn't beg any questions against the dualist—at least, not in the context of the present debate.

5 Conclusion

Let us take stock. First, we saw that physicalism admits of two basic flavors: extrinsic physicalism and intrinsic physicalism. The conceivability argument is held by dualists to refute physicalism. Contrary to this assertion, a number of authors have argued that *if* the conceivability argument is sound, then physicalism must be inconceivable (IN)—and this, as a matter of pure logic. I have offered that if we are sensitive to the flavor of physicalism under consideration, (IN) *is* a logical truth for intrinsic physicalism—that is, (INIP) is logically true—whereas (IN) is *not* a logical truth for extrinsic physicalism—that is, (INEP) is *not* a logical truth. In Sect. 2, I justified these claims, first by using the symmetric mirror argument to prove (INIP), and second by providing a counter-model to (INEP). In Sect. 3, I considered two pre-existing ways of augmenting (INEP) into a logical truth (INEP$_\psi$): a modal interpretation, which countenanced the axiom (5) and involved the conceivability of extrinsic physicalism; and an analytic interpretation, which countenanced the premise (A) and involved the conceivability of a physical situation. Both interpretations, I argued, failed to make (INEP) as problematic as (INIP). After all, the dualist is not committed to the (5) axiom, and she needs be committed to (A) only insofar as (INIP) is true—and no further. As such, (INIP) emerges as the serious threat to anti-physicalist dualism; intrinsic physicalism, in some sense, is the zombie situation's cleanest and most symmetric counterpart. Objections based on the entailment from intrinsic to extrinsic physicalism are themselves blocked, owing to a disjunctive argument: if the Kripkean premise is granted, so that the entailment holds, then any *uniquely* extrinsic argument *becomes* question-begging; if the Kripkean premise is *not* granted, then the entailment fails, and the objections are disarmed. I therefore submit that intrinsic reductive physicalism (and its *entailed*

extrinsic cousin, but *not* this alone), is the preferred doctrine for those physicalists sympathetic to the intuitions underlying the conceivability argument.

This result may be surprising (at least, I *hope* that it is). For it is a by-now familiar myth in the philosophy of mind that intrinsic reduction, in the form of the mind–body identity theory (Place 1956; Smart 1959 etc.), was summarily defeated by the mid-twentieth century. On the one hand, this was due to the plausibility of *multiple realization* (e.g., Putnam 1967; Fodor 1974; cf. Shapiro 2004), according to which *many different* sorts of physical properties could ground phenomenal ones. On the other hand, it was due to the modal arguments propounded in Kripke's *Naming and Necessity* (1980; cf. his 1971). If we indulge in some Whig history, both phenomena contributed to the reductivist's move away from intrinsic physicalism and toward *mere* extrinsic physicalism—especially in the form of metaphysical supervenience. But I do not think that multiple realization is a serious threat to intrinsic physicalism, especially under its broad physico-functional construal given here (cf. arguments in Jackson et al. 1982). For the identification of any phenomenal property with some *functional* one is obviously consistent with the multiple *physical* realization of the latter (and, via identity, of the former; cf. Lewis 1972). Furthermore, Kripke's argument is really just a variant of the conceivability argument, setting aside the entailed conclusion against extrinsic physicalism. As such, it is susceptible to (INIP), as argued above, and can be abandoned with some impunity. In addition to the apparent explanatory weakness of *mere* extrinsic physicalism (see Kim 1993), these considerations suggest that the myth is overblown. It pays to go intrinsic.

I want to close with two generalizations. The first is of interest to the philosophy of mind. For all I have said, intrinsic physicalism needn't be *strictly* limited to the identity theory (see VandenHombergh 2017 for a similar point). It suffices that the reduction relation be at least as strong as identity, in the sense that it supports the Kripkean (non-)necessities discussed above. Hence the broad moniker "intrinsic"; for all I can see, other non-extrinsically construed forms of physicalism might be equally as resilient to the conceivability argument [perhaps, e.g., grounding (Fine 2012), realization (Shoemaker 2007), etc.]. The second is of interest to modal epistemology, broadly construed. It strikes me that the discussion above needn't apply *strictly* to the problem of consciousness, but to *any* purported reduction in *any* other domain with modal stature. In this sense, it may be beneficial to adopt intrinsic reduction about, say, moral, mathematical, aesthetic, or other mysterious properties, regardless of the reducing base in which one might be interested. The strength of conceivability intuitions opposing any such reduction is, I suggest, inversely proportional to the logical strength of that reduction. While I certainly do not wish to live in a world of frivolous identities, they are for all the standard reasons of some serious, if occasional, use. We might not, however, have expected to find *more* of that use behind enemy lines—a good place to find it, if any.

References

Bayne, S. (1988). Kripke's Cartesian argument. *Philosophia, 18,* 265–270.

Bennett, K. (2007). Mental causation. *Philosophy Compass, 2*(2), 316–337.

Block, N. (1972). Troubles with functionalism. *Minnesota Studies in the Philosophy of Science, 9,* 261–325.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18,* 227–287.

Brandt, R., & Kim, J. (1967). The logic of the identity theory. *Journal of Philosophy, 64*(17), 515–537.

Brown, R. (2010). Deprioritizing the a priori arguments against physicalism. *Journal of Consciousness Studies, 17*(3–4), 47–69.

Campbell, D., Copeland, J., & Deng, Z.-R. (2017). The inconceivable popularity of conceivability arguments. *Philosophical Quarterly, 67*(267), 223–240.

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. Gendler & J. Hawthorne (Eds.), *Conceivability and possibility* (pp. 145–200). Oxford: Oxford University Press.

Chalmers, D. (2010). The two-dimensional argument against materialism. In D. Chalmers (Ed.), *The character of consciousness*. Oxford: Oxford University Press.

Chalmers, D., & Jackson, F. (2001). Conceptual analysis and reductive explanation. *Philosophical Review, 110*(3), 315–361.

Descartes, R. (1993). *Discourse on method; and, meditations on first philosophy*. Indianapolis: Hackett Pub. Co.

Fine, K. (2012). Guide to ground. In F. Correia & B. Schieder (Eds.), *Metaphysical grounding* (pp. 37–80). Cambridge: Cambridge University Press.

Fodor, J. (1974). Special sciences, or the disunity of science as a working hypothesis. *Synthese, 28*(2), 97–115.

Frankish, K. (2007). The anti-zombie argument. *Philosophical Quarterly, 57*(229), 650–666.

Garson, J. (2013). *Modal logic for philosophers*. Cambridge: Cambridge University Press.

Gertler, B. (2008). In defence of mind-body dualism. In J. Feinberg & R. Shafer-Landau (Eds.), *Reason & responsibility: Readings in some basic problems of philosophy* (pp. 285–297). Belmont: Thomson Wadsworth.

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

Jackson, F. (1998). *From metaphysics to ethics: A defence of conceptual analysis*. Oxford: Oxford University Press.

Jackson, F., Pargetter, R., & Prior, E. (1982). Functionalism and type-type identity theories. *Philosophical Studies, 42*(2), 209–225.

Kim, J. (1989). The myth of nonreductive materialism. *Proceedings and Addresses of the American Philosophical Association, 63*(3), 31–47.

Kim, J. (1993). *Supervenience and mind*. Cambridge: Cambridge University Press.

Kim, J. (2011). *Philosophy of mind* (3rd ed.). Boulder: Westview Press.

Kirk, R. (2005). *Zombies and consciousness*. Oxford: Oxford University Press.

Kripke, S. (1971). Identity and necessity. In M. Munitz (Ed.), *Identity and individuation* (pp. 135–164). New York: New York University Press.

Kripke, S. (1980). *Naming and necessity*. Cambridge: Harvard University Press.

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy, 50*(3), 249–258.

Marton, P. (1998). Zombies versus materialists: The battle for conceivability. *Southwest Philosophy Review, 14*(1), 131–138.

Naegle, C. (2012). The paradox of phenomenal judgment and causality. *Aporia, 22*(1), 11–19.

Nagel, E. (1961). *The structure of science: Problems in the logic of explanation*. San Diego: Hartcourt, Brace & World.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83,* 435–450.

Plantinga, A. (1974). *The nature of necessity*. Clarendon: Clarendon.

Putnam, H. (1967). Psychological predicates. In W. Capitan & D. Merrill (Eds.), *Art, mind, and religion* (pp. 37–48). Pittsburgh: University of Pittsburgh Press.

Rudd, A. (2000). Phenomenal judgment and mental causation. *Journal of Consciousness Studies, 7*(6), 53–66.

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–457.

Shapiro, L. (2004). *The mind incarnate*. Cambridge: MIT Press.

Shoemaker, S. (2007). *Physical realization*. Oxford: Oxford University Press.

Smart, J. J. C. (1959). Sensations and brain processes. *The Philosophical Review, 68*(2), 142.

Sober, E. (2015). *Ockham's Razors: A user's manual*. Cambridge: Cambridge University Press.

Sturgeon, S. (2000). *Matters of mind: Consciousness, reason, and nature*. Abingdon: Routledge.

VandenHombergh, J. (2017). Inconceivable physicalism. *Analysis, 77*(1), 116–125.