

Inconceivable physicalism

JONATHON VANDENHOMBERGH

1. *The setup*

Phenomenal facts are facts about the ways things feel or ‘what it is like’ to feel them (Nagel 1974)—facts witnessed by instances of phenomenal properties or states, like *searing pain* or *tingling orgasm* or *unendurable boredom*. It is hoped by some philosophers—perhaps fewer and fewer, these days—that phenomenal facts are somehow *reducible to* or *reductively explainable in terms of* physical facts, or, by way of some realization relation, the functional facts they make true (or both). This position is oftentimes called *reductive physicalism*.

The ‘two-dimensional conceivability argument’ (Chalmers 2010) purports to refute reductive physicalism, on the basis of both a priori reflection and the two-dimensional semantic content of our concepts or expressions for physical and phenomenal states of affairs. In its simplest form, the argument is as follows. Where P is the sum of actual, fundamental physical facts (and, if you’d like, indexical facts, and a closure clause specifying that nothing further is involved—‘that’s all’), and Q is any phenomenal fact (e.g., ‘I am in pain’):

- (1) It is conceivable that $P \ \& \ \sim Q$.
- (2) If it is conceivable that $P \ \& \ \sim Q$, then it is epistemically possible that $P \ \& \ \sim Q$.
- (3) If it is epistemically possible that $P \ \& \ \sim Q$, then it is metaphysically possible that $P \ \& \ \sim Q$.
- (4) Therefore, reductive physicalism (‘materialism’) is false.

The argument is valid, but the last step may not be obvious. Note only that, by (1)–(3), we may infer that

- (3.1) It is metaphysically possible that $P \ \& \ \sim Q$.

which is equivalent to

- (3.2) It is not metaphysically necessary that P implies Q.

And to the extent that reductive physicalism implies¹ the negation of (3.2), it will be automatically falsified.

1 ‘Entailment’ and ‘implication’ admit of multiple readings, especially in the context of the present debate (e.g., ‘a priori entailment,’ ‘logical implication,’ etc.). I will always treat unqualified ‘entailment’ or ‘implication’ as material implication in the classical propositional sense (i.e., as captured by ‘ \supset ,’ or, given the deduction theorem, ‘ \rightarrow ’). Thanks to an anonymous referee for pointing out this ambiguity.

There is a lot that can be—and has been—said about this argument (call it ‘2DA’). There is far more to it than meets the eye; so, far more to which one might object. To strategically unpack only a *bit* of it, let’s focus on premiss (3). In that premiss, we are asked to endorse a connection between the *epistemic* and *metaphysical* possibilities that $P \ \& \ \sim Q$. Given a single space of (metaphysically) possible worlds, this amounts to a connection between the *epistemic* and *metaphysical truth evaluation* of $P \ \& \ \sim Q$ (i.e., S is epistemically possible iff S is epistemically true—‘verified,’ as Chalmers (2011) puts it—in some possible world). Lastly, given that the truth conditions of logical connectives like ‘&’ and ‘~’ are invariant across both kinds of truth evaluation, this amounts to a connection between the epistemic and metaphysical truth evaluation of P and of Q . The connection itself is variously labelled *intensional coincidence* (Chalmers 2010; Nina-Rümelin 2006) or *semantic stability* (Balog 2009; Bealer 1996) or sometimes the tongue-twisting *non-twin-earthability* (Chalmers 2014). Whatever the name, it relies upon the modal notion that there are two distinct intensions for P (Q), one mapping from possible worlds *considered epistemically* or *as actual* to truth values (thus modeling epistemic truth evaluation), the other mapping from possible worlds *considered metaphysically* or *as counterfactual* to truth values (thus modeling metaphysical truth evaluation). P (Q) then has coinciding intensions—instantiates the relevant connection—just when, no matter the world w , both intensions assign w the same truth value. In what follows, I will grant this connection without argument, letting $e(\varphi)$ represent the epistemic intensional value of any statement φ , and letting ‘It is metaphysically necessary that $e(\varphi)$ if and only if φ ’ represent the claim that φ has coinciding intensions.

We can now recast 2DA more explicitly:

- (1) It is conceivable that $P \ \& \ \sim Q$.
- (2) If it is conceivable that $P \ \& \ \sim Q$, then it is epistemically possible that $P \ \& \ \sim Q$.
- (3) It is metaphysically necessary that $e(P)$ if and only if P , and likewise for Q .
 - (3.1) It is metaphysically possible that $P \ \& \ \sim Q$.
 - (3.2) It is not metaphysically necessary that P implies Q .
- (4) Therefore, reductive physicalism (‘materialism’) is false.

To be clear: by (1) and (2), we have it that $P \ \& \ \sim Q$ is epistemically possible. By the semantic picture sketched above, this means that $P \ \& \ \sim Q$ is epistemically true in some possible world. And *this* means that $e(P) \ \& \ \sim e(Q)$ is metaphysically true in some possible world. Then, by (3), we may infer that $P \ \& \ \sim Q$ is true in some possible world. Once again, therefore, reductive physicalism (in numerous forms) will be false.

2. *The argument*

Consider now two *names*, p and q , which respectively denote physical and phenomenal state tokens—e.g., *this pattern of nociceptor firing* and *that searing pain*.² Is it conceivably true that $p=q$? I think it is (and so do many others; Block and Stalnaker 1999; Lewis 1970, 1972; Smart 1959), but, without argument, let's see where this conceivability gets us. Surely, if the epistemic possibility of $P \ \& \ \sim Q$ is entailed by its conceivability, then the epistemic possibility of $p=q$ is likewise entailed: there is nothing about $P \ \& \ \sim Q$ *itself* that guarantees this entailment. Indeed, as Chalmers (2002) elsewhere goes through pains to show, conceivability of a certain precise form is supposed to be a *generally* reliable guide to epistemic possibility. So our assumption appears to get us the epistemic possibility of $p=q$. This, as before, comes to the possible epistemic truth of $p=q$, or the possible metaphysical truth of $e(p=q)$. Allowing us to treat '=' in its logical sense (as another operator, like '&' and '~'), we will have it that $e(p)=e(q)$ in some possible world. Once more, I see no reason to disadvantage identity if we do not disadvantage the connectives—there are robust, intensional and extensional truth conditions for the former, just as for the latter. Furthermore, it would once again appear to hold, by parity, that $e(p)=p$ and $e(q)=q$ are both metaphysically necessary, just as $e(P)$ iff P and $e(Q)$ iff Q are metaphysically necessary. In other words, if P and Q are both semantically stable—especially if this stability is parasitic on subsentential stability in P and Q —then p and q are stable as well. But then it is metaphysically possible that $p=q$; $p=q$ is true in some world, w . A well-known lesson of Kripke's—recapitulated by Chalmers—is that names are metaphysically rigid designators, denoting the same individual in all metaphysically possible worlds in which they exist. It is of course undeniable to all but the eliminativist or idealist that p and q refer in the actual world, and so their denotata exist here. But then they refer in the actual world as they do in w , given rigidity. If they are *identical* in w , then they refer to the *same individual there*. Therefore, they refer to the same individual in the actual world, and $p=q$ is *actually* (indeed, metaphysically and epistemically necessarily!) true. It appears that, given all of Chalmers's premisses but the first, something like the identity theory is true. This seems an odd consequence for (part of) an argument *against* reductive physicalism.

Actually, it's worse than an odd consequence. In what remains of this section, I'll prove that it entails a general and totally unsavory conditional:

2 The discussion extends, with some tinkering, to named state types and properties.

(Inconceivable Physicalism)

If n -many instances of 2DA are sound, then n -many instances of reductive physicalism are inconceivable.³

Here is how.

Given the definition of P, it is obvious that P will entail certain facts, such as that ‘Some particular pattern of nociceptor firing exists,’ symbolized as ‘ $\exists x x = p$,’⁴ and letting ‘p’ denote the relevant firing pattern.⁵ Now, the choice of Q is arbitrary—so long as it is a phenomenal fact, it counts, as reductivism is a global thesis relating *any* phenomenal fact to the total physical picture of the world. So, let Q = ‘Some searing pain exists,’ symbolized similarly as ‘ $\exists x x = q$.’ In that case, the following represents a uniformly and admissibly interpreted instance of 2DA, foregoing, at least momentarily, the sub-conclusions in lines (3.2) and (4):

- (1) It is conceivable that $P \ \& \ \sim \exists x x = q$.
- (2) If it is conceivable that $P \ \& \ \sim \exists x x = q$, then it is epistemically possible that $P \ \& \ \sim \exists x x = q$.
- (3) P and $\exists x x = q$ are semantically stable.

(3.1) It is metaphysically possible that $P \ \& \ \sim \exists x x = q$.

But P entails $\exists x x = p$, and this entailment holds necessarily, since P just *is* the conjunction of all physical facts (including $\exists x x = p$). Thus it follows that

(3.3) It is metaphysically possible that $\exists x x = p \ \& \ \sim \exists x x = q$.

Yet (3.3) is inconsistent with $p = q$, given the condition of rigidity. For assume that the identity theory is true:

(3.4) $p = q$.

Rigidity entails that

(3.5) It is metaphysically necessary that $p = q$.

But by (3.3), we have it that

(3.6) It is metaphysically possible that $\exists x x = p$.

3 Note that this is simply a formalization of the core thesis found in the abstract: that the soundness of 2DA entails the inconceivability of a wide swath of reductive theses.

4 The symbolism is a bit cumbersome, but intentional: first, it allows me to forego philosophical controversy over an independent ‘existence’ predicate (I use only the quantifier and identity); second, it makes explicit the relation to ordinary identities like $p = q$ (which of course generalizes to $\exists x x = q$). This relation is exploited in the arguments below.

5 Of course, this isn’t expressed in the language of fundamental microphysics. But given our assumption about the reduction of non-physical sciences to physical ones, it will plausibly count. At any rate, the example is immaterial—just pick whatever instance of p you’d like, so long as it is appropriate for a putative identity $p = q$.

Thus there is some world w in which $\exists x x=p$. But, by (3.5), w is also a world in which $\exists x x=q$; after all, if there is a p , and p is necessarily identical to q , then there is a q . Yet we *also* see, once more from (3.3), that w is a world in which $\sim\exists x x=q$. So, both $\exists x x=q$ and $\sim\exists x x=q$ hold in w , which is contradictory. It follows that our assumption, $p=q$, is false.

But then, granting all the other premisses in 2DA, $p=q$ is conceivable only if $P \ \& \ \sim\exists x x=q$ isn't, where, once again, P is a massive conjunction entailing $\exists x x=p$ and $\exists x x=q$ is an instance of Q . For consider: if $p=q$ is conceivable, then, as we showed, $p=q$ is true, *simpliciter* (this was our initial 'odd' consequence). If it is *also* conceivable that $P \ \& \ \sim\exists x x=q$, then $p=q$ must be false, as above. So, if $p=q$ is conceivable, $P \ \& \ \sim\exists x x=q$ can't be, just as suggested. Of course, this argument will apply for any existent physical state p entailed by P , and for any existent q . So we have it that

(3.7) For any existing physical and phenomenal states p and q , $p=q$ is conceivable only if $P \ \& \ \sim\exists x x=q$ is not.

(3.7) gets us about halfway to the unsavory (Inconceivable Physicalism), and it follows straightforwardly from the initial premisses.

Let us now suppose that there are n -many actually existing phenomenal states:

$\exists x x = q_1$
 $\exists x x = q_2$
 ...
 $\exists x x = q_n$,

and n -many instances of 2DA opposing the identity of each q_i with some p_i , the existence of which is, just as before, straightforwardly entailed by P :

2DA-1 (refutes $p_1 = q_1$)
 2DA-2 (refutes $p_2 = q_2$)
 ...
 2DA- n (refutes $p_n = q_n$)

But then let us instantiate (3.7) for each of these p s and q s, yielding:

$p_1 = q_1$ is conceivable only if $P \ \& \ \sim\exists x x = q_1$ is not.
 $p_1 = q_2$ is conceivable only if $P \ \& \ \sim\exists x x = q_2$ is not.
 ...
 $p_n = q_n$ is conceivable only if $P \ \& \ \sim\exists x x = q_n$ is not.

It is obvious that the consequent of each conditional above is just the negation of premiss 1 for its respective instance of 2DA. Therefore, in each case, the consequent entails that *some* 2DA is unsound, since its first premiss is false:

$p_1 = q_1$ is conceivable only if 2DA-1 is unsound.

$p_1 = q_2$ is conceivable only if 2DA-2 is unsound.

...

$p_n = q_n$ is conceivable only if 2DA- n is unsound.

More generally, we have it that:

(3.8) For n -many actually existing physical and phenomenal states p and q , $p = q$ is conceivable only if n -many instances of 2DA are unsound.

But, by contraposition, (3.8) is equivalent to:

(3.9) For n -many actually existing physical and phenomenal states p and q , if n -many instances of 2DA are sound, then $p = q$ is inconceivable.

So that a *necessary condition* on n -many of these 2DA instances is the *inconceivability of n -many instances of the identity theory*. If the identity theory counts as a robust form of reductive physicalism—what could be more robust?—then (3.9) gives us the promised conditional:

(Inconceivable Physicalism)

If n -many instances of 2DA are sound, then n -many instances of reductive physicalism are inconceivable.

(Inconceivable Physicalism), I submit, vindicates that 2DA is worse than a mere oddity. Furthermore, it follows merely from an application of two-dimensionally valid inferences, the definitions of P and Q, and the general principles (e.g., conceivability entails epistemic possibility) which ground Chalmers's original premisses.

3. *The upshot*

The takeaway is rhetorical, for both parties to the dispute. If you are an identity theoretic reductivist, (Inconceivable Physicalism) will make explicit the question-begging nature of 2DA: how could it be rational for you to entertain an argument against one of your beliefs, if that argument requires you to find that belief *inconceivable*? If you are an anti-reductivist, (Inconceivable Physicalism) will make explicit the irrelevance of 2DA: how could it be rational to argue against an inconceivable position? Perhaps an anti-reductivist will reply—as, in a somewhat distinct vein, Chalmers has (2010: § 8)—that there is here a bullet worth biting: perhaps the identity theory *isn't* conceivable, after all. There are some rules for playing *that*

game, supplied by Chalmers's (2002) taxonomy of conceivability, and I think that we could in fact victoriously play it.⁶ But as this is a note, I will leave it to the reader to decide the adequacy of this anti-reductive reply.

This objection also has its analogs. Marton (1999) and Sturgeon (2000) both rebut 2DA by noting that the conceivability of reductive physicalism is inconsistent with the conceivability of the zombie world, embodied in that of $P \ \& \ \sim Q$. Yet these claims (as the authors note) depend crucially on two assumptions⁷ (assumptions which, incidentally, Chalmers is quick to point out). To see them, let reductive physicalism *amount to* $\Box(P \supset Q)$; the mere supervenience claim that 'it is metaphysically necessary that P entails Q'. We could then apply the premisses of 2DA, much as we did for $p = q$, as follows:

- (5) It is conceivable that $\Box(P \supset Q)$.
- (6) If it is conceivable that $\Box(P \supset Q)$, then it is epistemically possible that $\Box(P \supset Q)$.
- (7) It is metaphysically necessary that $e(P)$ if and only if P , and likewise for Q .

Difficulties accrue after (7). From (5) and (6), we may of course infer that $\Box(P \supset Q)$ is epistemically possible. And so $\Box(P \supset Q)$ is verified or epistemically true in some possible world, w . But how do we proceed from here? How do we apply the stability claim in line (7)? We might suppose that any $\Box(\varphi)$ is epistemically true in some world w if and only if φ is true in all epistemically possible worlds accessible from w ; consequently, iff $e(\varphi)$ is metaphysically true (φ is epistemically true) in all possible worlds accessible from w . Granting this assumption and the truth functionality of the epistemic intension, $e(P) \supset e(Q)$ will then be metaphysically true in all worlds accessible from w . Yet premiss (7) cannot get us to $P \supset Q$ in all worlds accessible from w , as hoped. For that, (7) must *itself* be metaphysically necessary: intensional

6 So as not to tease: Chalmers offers that it may be 'prima facie negatively conceivable that materialism is true about consciousness, but it is not obviously conceivable in any stronger sense' (2010). Any φ is *prima facie* negatively conceivable roughly when φ cannot be cursorily ruled out on a priori grounds (Chalmers 2002). Yet one might suggest that, insofar as p and q have coinciding intensions, the *prima facie* negative conceivability of $p = q$ is *insensitive* to the epistemic and psychological constraints imposed by the qualifiers 'prima facie' and 'negative'—after all, $p = q$ will automatically be conceivably *secondarily true* (or *metaphysically true*), and secondary truth is insensitive to epistemic and psychological constraints. Yet secondary conceivability is arguably a robust guide to metaphysical possibility, and so (Inconceivable Physicalism) reemerges. Compare Chalmers's more careful discussion of secondary conceivability in his 2002; see also Kripke 1980.

7 Marton makes this assumption explicit. Sturgeon does not, but he endorses a logic for the metaphysical modality equivalent to S4—semantically, a logic under which worlds are transitively accessible (see p. 116 of his 2000). This is not quite S5, of course, but my discussion straightforwardly applies to both S4 and S5. For simplicity, I will make reference only to the latter.

coincidence must hold in all worlds *accessible from* w —a world which might *differ* from the one at which (7) is evaluated as true.⁸ And, even granting that (7) is necessary, it will *at most* follow that it is metaphysically *possible* that $\Box(P \supset Q)$; it is *possible* that reductive physicalism is true. The original conclusion of 2DA will not be inconsistent with this possibility, so there will be no means of exploiting the inconsistency as in (Inconceivable Physicalism).

We can avoid this roadblock if we endorse an underlying *universal* accessibility relation between possible worlds: a relation syntactically circumscribed by the logic S5, according to which (*inter alia*) $\Box\phi \supset \Box\Box\phi$ and $\Diamond\Box\phi \supset \Box\phi$ are both valid schemata (and where ‘ \Diamond ’ represents metaphysical possibility). Fair enough: but (Inconceivable Physicalism) doesn’t need S5, and it doesn’t need anything else, except for what Chalmers has given us. If we value parsimony, it is the desirable thesis.

Yet perhaps the anti-reductivist is wondering why it matters. After all, there are *other* forms of reductive physicalism which appear to be falsified by 2DA, and yet which do not lead to (Inconceivable Physicalism). Why wouldn’t their falsification impugn reductive physicalism more generally, by controverting non-identity theoretic physicalisms—even if the identity theory itself remains safe? First, consider the mere supervenience forms of physicalism, as directly attacked by 2DA and defended by Marton and Sturgeon. These are indeed falsified by 2DA, and they do not lead to (Inconceivable Physicalism). Yet there is simply no reason the physicalist *needs* to endorse mere supervenience physicalisms, especially when identity does the reductive job in a far more explanatorily effective way (cf. Kim 1993’s related complaints about the explanatory limits of mere supervenience). Second, consider forms of reductive physicalism which are neither identity theoretic nor merely supervenient—for instance, those which implicate deeper relations of grounding (e.g., Fine 2012), or realization (e.g., Shoemaker 2007), or what have you. Aren’t *these* falsified by 2DA, and, if so, isn’t physicalism once more impugned? Not so: to the extent that these deeper relations are *intrinsic*—they are relations for which mere metaphysical necessitation is insufficient—it would seem that any 2DA instance opposed to them would once again lead to some nearby versions of (Inconceivable Physicalism). More simply: their conceivability will once again entail their truth, and, *a fortiori*, the 2DA instances opposed to them will entail their inconceivability.⁹ Defenders of 2DA are thus trapped between states of explanatory irrelevance and (Inconceivable Physicalism)—no doubt further forms of reductivism, if they exist, would be sorted into one of these two

8 Thanks to an anonymous referee for pressing me on the details of this section.

9 More precisely: to the extent that they are intrinsic, these relations will not obviously require S5-style accessibility (much like identity), and so their conceivable truth will generate similar (parsimonious) versions of (Inconceivable Physicalism). Thanks to an anonymous referee for this interesting suggestion.

camps. As such, the anti-reductivist ought to sidestep both Scylla and Charybdis entirely; they ought to sail back to the waters of less contentious argumentation. Better still, if the spirit of (Inconceivable Physicalism) generalizes to other anti-reductive intuitions and arguments, they might just join us physicalists on the shoreline.¹⁰

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

References

- Balog, K. 2009. Phenomenal concepts. In *The Oxford Handbook of the Philosophy of Mind*, eds. B. McLaughlin, A. Beckermann, and S. Walter, 292–312. Oxford: Oxford University Press.
- Bealer, G. 1996. On the possibility of philosophical knowledge. *Philosophical Perspectives* 10: 1–34.
- Block, N. and Stalnaker, R. 1999. Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review* 108: 1–46.
- Chalmers, D. 2002. Does conceivability entail possibility? In *Conceivability and Possibility*, eds. T. Gendler and J. Hawthorne, 145–200. Oxford: Oxford University Press.
- Chalmers, D. 2010. The two dimensional argument against materialism. In *The Character of Consciousness*, ed. D. Chalmers, 141–205. Oxford: Oxford University Press.
- Chalmers, D. 2011. The nature of epistemic space. In *Epistemic Modality*, eds. A. Egan and B. Weatherson. Oxford: Oxford University Press.
- Chalmers, D. 2014. *Constructing the World*. Oxford: Oxford University Press.
- Fine, K. 2012. Guide to ground. In *Metaphysical Grounding*, eds. F. Correia and B. Schnieder, 37–80. Cambridge: Cambridge University Press.
- Kim, J. 1993. *Supervenience and Mind*. Cambridge: Cambridge University Press.
- Kripke, S. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.
- Lewis, D. 1970. An argument for the identity theory. *Journal of Philosophy* 63: 17–25.
- Lewis, D. 1972. Psychophysical and theoretical identifications. *Australasian Journal of Philosophy* 50: 249–58.
- Marton, P. 1999. Zombies vs. materialists: the battle for conceivability. *Southwest Philosophy Review* 14: 131–8.
- Nagel, T. 1974. What is it like to be a bat? *Philosophical Review* 83: 435–50.

10 Thanks to audiences at the University of Houston and Stockholm University for feedback and helpful suggestions on earlier drafts of this paper. I am particularly indebted to David Chalmers, Chris Frugé, James Garson, Bredo Johnsen, Meryem Keskin, Kyle Landrum, Pavel Nitchovski, Josh Weisberg, and anonymous referees from *Analysis*. This paper is dedicated to my parents and to Meryem Keskin, for their constant love and encouragement.

- Nina-Rümelin, M. 2007. Grasping phenomenal properties. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, eds. T. Alter and S. Walter. New York: Oxford University Press.
- Shoemaker, S. 2007. *Physical Realization*. Oxford: Oxford University Press.
- Smart, J.J.C. 1959. Sensations and brain processes. *Philosophical Review* 68: 141–56.
- Sturgeon, S. 2000. *Matters of Mind: Consciousness, Reason, and Nature*. London: Routledge.