Philos Stud (2016) 173:2353–2375
DOI 10.1007/s11098-015-0618-7

# How to befriend zombies: a guide for physicalists

**Bradford Saad**[1]

Published online: 29 December 2015


**Abstract** Though not myself a physicalist, I develop a new argument against antiphysicalist positions that are motivated by zombie arguments. I first identify four general features of phenomenal states that are candidates for non-physical types; these are used to generate different types of zombie. I distinguish two antiphysicalist positions: strict dualism, which posits exactly one general non-physical type, and pluralism, which posits more than one such type. It turns out that zombie arguments threaten strict dualism and some pluralist positions as much as they threaten physicalism—indeed, more so, since such positions need zombies to motivate them as alternatives to physicalism—and that the only pluralist position that escapes zombie arguments has a radically inflated ontology.

**Keywords** Consciousness · Physicalism · Zombies · Dualism · Conceivability

## 1 Introduction

Zombies—(roughly) unconscious duplicates of actual conscious creatures—are widely viewed as foes of physicalism, and therefore as friends of antiphysicalism. After all, the possibility of zombies would imply the falsity of physicalism. Though not myself a physicalist, the aim of this paper is to show that, contrary to the noted view, zombies can, to the detriment of antiphysicalist positions, be enlisted in the service of physicalism.

The positions I discuss agree that the fundamental truths (about our world) ground all truths (about our world). They disagree about which general types of truths are the (irreducible) ones whose fundamental instances are needed to ground

✉ Bradford Saad
bradsaad@utexas.edu

[1] The Philosophy Department, University of Texas at Austin, Austin, TX, USA

all truths.[1] *Physicalists* claim that the fundamental physical truths are the only ones needed to ground all truths.[2] *Antiphysicalists* deny this, claiming instead that the fundamental physical truths, along with the fundamental truths of at least one other general type, are needed to ground all truths. Antiphysicalism comes in two relevant forms: *strict dualism* and *pluralism*. Strict dualists and pluralists disagree about how many general types of truth are needed to ground all truths. According to the strict dualist, *exactly two* types are needed (one of which is the physical, the other of which is some general non-physical type).[3] According to the pluralist, *more than two* types are needed (one type of which is the physical).

The core argument, which I will be developing on behalf of the physicalist but not endorsing, can be stated as follows:

1. If strict dualism and pluralism should be rejected, antiphysicalism should be rejected (and physicalism accepted).
2. Strict dualism should be rejected.
3. Pluralism should be rejected.
4. So, antiphysicalism should be rejected (and physicalism accepted).

The non-parenthetical reading of 1 is trivially true, as strict dualism and pluralism by definition exhaust antiphysicalist positions. On the other hand, since strict dualism and pluralism do not quite exhaust non-physicalist alternatives (on which the fundamental truths ground all truths)—specifically, they do not encompass idealist positions that take non-physical truths to be the only ones needed to ground all truths—the parenthetical reading of 1 is not quite trivially true. Still, given that the fundamental truths ground all truths and that such idealist positions are generally not regarded as live possibilities in contemporary philosophy of mind, 1 should be relatively uncontroversial even on the parenthetical reading.[4] I will not be concerned to defend the parenthetical reading of 1 by arguing against positions

[1] I assume that A grounds B only if A metaphysically necessitates B. For concision, I will generally leave the about-our-world restriction and irreducibility requirement implicit.

I think my argument can be extended to views that deny that all truths are grounded in fundamental truths on the ground that for every level, there exists a more fundamental level by using (something like) Montero's (2006, p. 187) technique for defining 'physicalism' so as to allow for its truth in infinitely decomposable worlds. Since I cannot attempt that extension here, I set such views aside.

[2] Of course, formulating physicalism (and its rivals) is, in general, a tricky business. Formulating physicalism in terms of grounding is no exception. Though imperfect, the formulation in the main text suffices for my purposes. I adopt it primarily for convenience. Once I have presented my argumentative strategy using this formulation, applying the strategy with other common formulations of physicalism should be fairly straightforward. For discussion of how to formulate physicalism in terms of grounding, see Bennett (2011), Dasgupta (2014), and Goff (ms). For a formulation of physicalism about consciousness in terms of grounding, see Chalmers (2015, p. 248).

[3] I label this position 'strict dualism' rather than just 'dualism' because dualism is rarely characterized as being committed to there being *exactly* (as opposed to at least) two irreducible general types of truths (which is not to say that it is generally characterized as not being so committed).

[4] Some panpsychist positions that take fundamental (proto)phenomenal qualities to pervade the external world are antiphysicalist, not idealist. For example, a panpsychist who takes the physical to consist in structure and dynamics and thinks that fundamental structure and dynamics are necessary (albeit not sufficient) to ground all truths, would be an antiphysicalist, not an idealist.

that reject both physicalism and antiphysicalism. In any event, proponents of such positions may read the argument non-parenthetically as an argument against antiphysicalism, while those content to set aside such positions may read it parenthetically as an argument against antiphysicalism and for physicalism.

Most of my discussion focuses on defending 2. My defense of 2 takes advantage of recent observations by Chalmers (forthcoming, §4.3, 2015, p. 257, §6, §7) and Goff (2009) in their discussions of panpsychism: there are different types of zombie corresponding to different explanatory gaps in the vicinity of the phenomenal–physical explanatory gap. I argue that (pairs of) these zombies can be marshalled against strict dualists. That is, zombie arguments against physicalism turn out to have analogs that threaten strict dualism. As we will see, this forces an unwelcome decision on strict dualists: either (1) reject the possibility of zombies wholesale, thereby giving up one of the most powerful considerations (i.e. the zombie argument) against physicalism or (2) make the alliance with zombies selective by accepting the possibility of one type of zombie while rejecting the possibility of every other. Neither option is acceptable: (1) comes at the price of leaving strict dualism unmotivated and physicalism unthreatened, while (2) generates the obstacle of providing a principled basis for preferential modal treatment of different types of zombie. Once the inevitability of this decision is recognized, strict dualism ceases to be an acceptable position.

(For vividness, I allow zombies to loom large in my discussion. However, once I show how zombie arguments, which are typically used against physicalism, can be used against strict dualism, it should be fairly obvious how to modify other standard arguments against physicalism to target strict dualism. I will have more to say about this in Sect. 4.)

Antiphysicalists who wish to retain the service of zombies, then, must forgo strict dualism for pluralism. However, I argue that the only way to adopt a motivated form of pluralism is to adopt a *radical* form of pluralism that posits *significantly more* than two general types of truths for the purposes of grounding all others. Such ontological inflation, the physicalist may maintain, is unacceptable. Thus, on pain of unacceptable ontological inflation, radical pluralism, and hence pluralism must be rejected, or so goes the defense of 3 I offer on behalf of the physicalist. Most of the defense of 3 consists in demonstrating just how inflationary a motivated pluralist position needs to be, as opposed to arguing for the badness of ontological inflation.

Having defended premises 1–3, I will have charted a new route to the conclusion 4, i.e. that antiphysicalism should be rejected (and physicalism accepted). While I am not ultimately convinced by the argument—I do not accept 3, but nor do I regard its acceptance as irrational—I find it more convincing than any extant argument against antiphysicalism (and for physicalism) of which I am aware. The upshot will be that physicalists can use zombies to disarm and demotivate zombie attacks from ontologically moderate antiphysicalists, holding a powerful parsimony objection in reserve should such theorists retreat to the ontologically inflationary ranks of the remaining antiphysicalist position, namely radical pluralism. Therein emerges the possibility of physicalist-zombie friendship.

The paper is structured as follows. In Sect. 2, I characterize different features of phenomenal states that can be used to generate different types of zombie. In Sect. 3,

I introduce zombies and zombie arguments and show how they can be used against strict dualism, thereby justifying premise 2. In Sect. 4, I give reasons for thinking that just as zombie arguments against physicalism have analogs that threaten strict dualism, so too do other arguments against physicalism have analogs that threaten strict dualism. In Sect. 5, I develop an argument against pluralism to support premise 3. In Sect. 6, I characterize the dialectical import of my arguments.

## 2 Explanatory gaps galore

In this section, I use the explanatory gap framework as a presentational vehicle for distinguishing four features of phenomenal states that I use to generate different types of zombie and zombie arguments in the next section.

The traditional explanatory gap consists in the fact that phenomenal states do not seem to be explainable in terms of physical states [see Levine (1983)]. Phenomenal states are states that there is *something it is like to be in*.[5] In what follows, I distinguish four features of phenomenal states that individually resist explanation.[6] The point of presenting these explanatory gaps is *not* to argue for the non-physicality of these features. Rather, I use the explanatory gap framework as a familiar device to help the reader acquire a grip on exactly which features of phenomenal states are under discussion.

First, phenomenal states involve *qualities*. For example, when you have a visual experience as of a blue ball, you seem to experience a color quality extended over

[5] I will not attempt to characterize the physical, except to say that (a) my discussion is neutral between the standard characterizations of the physical and could be run in terms of any of them, and (b) my preferred characterization, cashes physicality out in terms of structure and dynamics (Chalmers 2010, p. 120).

[6] I focus on four features of phenomenal states that pose explanatory gaps, but there are others. For example, there is arguably a gap between *phenomenal* (spatial, temporal, mereological) *structure* and *physical structure*. [Cf. Chalmers (forthcoming, §3, §4.7, §8).] For reasons I hope to discuss on another occasion, I believe this gap cannot serve the same role in motivating conceivability arguments as the ones I discuss in the main text. Pan(proto)psychist views also face a number of explanatory gaps, such as those that arise between *macro-subjects* and *micro-subjects*, *macro-qualities* and *micro-qualities*, and *macro-experiences* and *micro-experiences*. [See Chalmers (forthcoming 2015) and Goff (2009).] Though some gaps that pan(proto)psychism faces are not relevant for my purposes, some gaps that I do discuss have been raised for pan(proto)psychism. Another gap resides between types and individuals. Even if we could explain why a physical type's instantiation necessitated a phenomenal type's instantiation, there would remain a question as to why an instance of that physical type co-obtains with a particular instance of that phenomenal type, as opposed to another instance of that type. I set this gap aside because it is a special case of what we might call the *haecceity–non-haecceity gap*, which arises in some form regardless of one's account of consciousness [cf. Lewis (2009, pp. 209–210, fn15)]. *Intentional* features of phenomenal states arguably pose a further gap. [See Goff (2012) and Pautz (2009).] However, delving into the metaphysics of intentionality would take us too far astray. I also set aside the *subjective–objective(/physical)* gap [e.g. see Nagel (1974, 1986)] because I have no clear notion of it that differs from notions of other explanatory gaps. Finally, there is arguably a gap between the *normative properties of experience* (and mentality more generally) and the *physical*. [See Lee (2013) and Sosa (2011).] However, the (alleged) a priori metaphysical supervenience of the normative (e.g. on the natural) poses an obstacle to using the normative in zombie arguments of the sort I offer below; see Yetter-Chappell and Chappell (2013, fn11).

the surface of an external object. We need not settle whether the quality in question is in fact instantiated by an external object—as it would be on a direct realist theory of perception, or not, as, for example, a projectivist theory would claim—to agree that such qualities are constitutively involved in phenomenal states. Assuming that such qualities are not fundamental physical properties, there is a gap between the physical and such qualities, as there does not seem to be a way of explaining the latter in terms of the former. Call this the *qualitative-physical (explanatory) gap*.

Second, phenomenal states involve *awareness*.[7] Consider (what I take to be) the commonsense view according to which the world contains a multitude of objects that have awareness-independent intrinsic non-dispositional color qualities, properties that we sometimes experience objects as having. On this commonsense view, objects can, and often do, instantiate color qualities without anyone experiencing the objects or their qualities. For example, carrots would, on this view, instantiate the quality that I call 'orange' not only when an observer has an experience of the object's being orange, but also when no one is having an experience of the object's being orange. [See Coleman (2012, p. 153) for similar points.]

I myself think that the commonsense view is probably false, but, more importantly, I hold that it is conceivably true. In conceiving of its truth, we coherently represent scenarios in which qualities exist independently of anyone's awareness of them (e.g. we can conceive of carrots being qualitative orange whether or not anyone is experiencing them).

The type of awareness in question seems to consist in the immediate, non-inferential apprehension of qualities.[8] On one attractive conception, awareness is a relation between a subject and a quality. However, this conception is not mandatory. Awareness might instead be understood as a property sometimes instantiated with/by qualities. The latter conception might be adopted, for example, by someone who is a realist about awareness and an eliminativist about subjects. I remain neutral between the relational and property conceptions of awareness.[9]

Assuming that awareness cannot be taken as a fundamental physical primitive, there is an explanatory gap between the physical and awareness. One way to argue for this gap is by appealing to the qualitative-physical gap. For awareness is awareness *of qualities*, and there is a qualitative-physical gap. Since a physical explanation of awareness would have to explain qualities, the existence of a qualitative-physical gap implies the existence of an *awareness-physical gap*.

[7] 'Acquaintance' might work just as well as a term for picking out this phenomenon as 'awareness'. Both terms have connotations (access consciousness and representational content in the case of 'awareness' and, in the case of 'acquaintance' sense data and foundational epistemology) that are not built into the notion I am using. See Pautz (2009) for argument that we bear an irreducible "consciousness relation" to the ("sensible") properties that determine phenomenal character.

[8] A tricky issue here that I will mention only to set aside: there is an intuition to the effect that the type of awareness we have of qualities is also constitutive of states—such as cognition, moods, and emotions—that do not (obviously) involve qualities. However, it is difficult to say what the immediate objects of such awareness are if not qualities. To avoid having to address this issue in what follows, I stipulate that 'awareness' is to be understood narrowly as the awareness of qualities.

[9] Cf. Chalmers (2015, §7, forthcoming, §6.4).

A natural question here is whether the awareness-physical gap is merely a manifestation of the qualitative-physical gap, or, if it instead poses a further explanatory lacuna. A moment's reflection suggests that there is indeed a further gap. For, given any distribution of physical and qualitative properties, there seems to be a further, open question about which qualities (if any) are objects of awareness. So, there is an *awareness-(qualitative-physical) gap*—i.e. a gap between awareness, on the one hand, and the physical and qualitative on the other. [See Chalmers (2015, pp. 273–274); cf. Nagel (1974, fn10).]

Third, some phenomenal states are *unified*, in that they are composite states of awareness, consisting of component states of awareness. For example, when (under normal circumstances) you taste coffee and smell basil at the same time, you instantiate one phenomenal state that consists in the awareness of a gustatory quality, a second that consists in the awareness of an olfactory quality, and a third, composed of the first two, that consists of your being aware of both a gustatory quality *and* an olfactory quality. (Compare: suppose you taste coffee but don't smell basil and a friend smells basil but doesn't taste coffee; in that case, between the two of you, there would be awareness of the noted gustatory quality and awareness of the noted olfactory quality, but no awareness of both.)[10]

Since phenomenal unity, as I understand it, is a relation between states of awareness of qualities, the existence of a unity-physical gap follows immediately from the existence of the quality-physical gap and from the existence of an awareness-physical gap. As before, there is a question as to whether the unity-physical gap is merely a manifestation of these gaps, or if it is instead a further gap in its own right. Again, there is arguably a further gap. For, in addition to the physical truths not a priori entailing the unity truths, there also seems to be a failure of a priori entailment from the conjunction of the physical, qualitative, and awareness truths to the unity truths. So, there is also a *unity-(qualitative-awareness-physical) gap*.

Finally, qualities, awareness of qualities, and unified awareness of qualities are all *possessed* by entities. That this is a further feature of phenomenal states can be discerned by noting that, for example, substance physicalists and substance dualists may agree on the nature of a unified phenomenal state, while disagreeing about what (type of) substance has that state. The substance physicalist may claim that such states are instantiated by physical brains, while the substance dualist may claim that such states are instantiated by non-physical minds. Similarly, a direct realist and a projectivist might agree that qualities are constitutive of visual experience while disagreeing about whether such qualities are instantiated by minds alone or instead by both minds and external bodies. In such cases, there is agreement about the existence of (components of) unified phenomenal states and disagreement about what possesses those states.

Is the *possession-physical gap* merely a manifestation of the gaps between the qualitative, awareness, and unity features of the phenomenal, on the one hand, and

[10] Cf. (Chalmers forthcoming, p. 5). For a good discussion of phenomenal unity, see Bayne and Chalmers (2003).

the physical, on the other? The answer is plausibly "no". For it seems that an ideally rational agent could know all the physical truths and all the truths about the distribution of qualitative, awareness, and unity properties without thereby being in a position to determine which entities possess unified phenomenal states (or their components).

Admittedly, some entities may be ruled out a priori as candidate bearers of unified phenomenal states. For example, on the supposition that qualitative properties are only instantiated outside of space, there is plausibly a priori warrant for concluding that they are not possessed by entities that exist wholly in space. However, the a priori eliminability of some candidate possessors is a far cry from an explanation of all the possession truths in terms of the physical, qualitative, awareness, and unity truths. Indeed, no such explanation suggests itself. So, there is a *possession-(unity-qualitative-awareness-physical) gap*.

It is worth noting that this gap might be partially bridged by, for example, positing or explaining the possession truths about qualities and/or awareness. Plausibly, there would remain a gap between (1) the physical truths, the truths about the distribution of qualities, awareness, and unity, and the truths about the possession of qualities and/or awareness and (2) truths about which entities possess unified phenomenal states.

I will use these features of phenomenal states in Sect. 3 to generate zombies and zombie arguments.

## 3 Zombies against strict dualism

### 3.1 Zombies

Zombies are standardly understood as phenomenally unconscious physical duplicates of phenomenally conscious creatures in the actual world. Assuming, as I do, that you are conscious, your zombie counterpart would, on this usage, be just like you physically, but there would be nothing it is like to be it. My usage slightly expands this standard usage in a couple of ways. A distinction is sometimes made between (a) zombies, understood as unconscious (physical) duplicates of conscious beings, (b) partial zombies, which are duplicates of conscious creatures except that they lack some of the phenomenal properties of their actual counterparts, and (c) inverts, which are duplicates of conscious creatures except that they have different phenomenal states instead of the ones had by their actual counterparts.

The possibility of any of (a)–(c) would falsify physicalism. So, any of these creatures can serve the zombie role in the zombie argument against physicalism. In light of this, I extend the application of 'zombie' to the creatures usually called 'inverts' and 'partial zombies'. I will also use the above explanatory gaps to distinguish several types of zombie that are not usually distinguished.

Talk about zombies is shorthand for talk about zombie scenarios, i.e. conceivable worlds that are duplicates (in some respects but not others) of the actual world.

Implementing zombie arguments against strict dualism requires distinguishing several types of zombie, arguing for their conceivability, and linking their

conceivability with their possibility. To that end, it will be useful to adopt the following notation.

***P***—the (sentential) conjunction of all physical truths (about our world).
***Q***—the conjunction of all qualitative truths.
***A***—the conjunction of all truths about which qualitative entities are objects of awareness.[11]
***U***—the conjunction of all truths about phenomenal unity.
***S***—the conjunction of all truths about which entities possess qualities, awareness, and phenomenally unified states.

I will understand strings of these letters as conjunctions (e.g. *PQ* is the conjunction of all physical and qualitative truths), and will use '$\sim$' to express negation. Strictly speaking, the statement of fully rigorous zombie arguments requires augmenting and restricting the truths figuring in *PQAUS* so as to avoid a number of technical problems.[12] However, to keep the discussion tractable, I ignore these problems and stick with simpler, albeit less rigorous, formulations.

### 3.2 A zombie argument against physicalism

In preparation for stating zombie arguments against strict dualism, we can, using this terminology, state a(n improved version of a standard) zombie argument against physicalism:

**P1**. $P \sim (QAUS)$ is conceivable.
**P2**. If $P \sim (QAUS)$ is conceivable, then $P \sim (QAUS)$ is possible.
**P3**. If $P \sim (QAUS)$ is possible, then physicalism is false.
**Conclusion**: So, physicalism is false.

P1 asserts the conceivability of all the physical truths obtaining without all the qualitative, awareness, unity, and possession truths obtaining. Though my arguments could be reasonably read with a number of different notions of conceivability, the one I will have in mind is *ideal positive conceivability*: a

[11] The gloss, which I will sometimes use, of ***A*** as the conjunction of awareness truths is potentially misleading. For on a one way of understanding the gloss, *A*'s conjuncts include truths about unified awareness (e.g. that the gustatory quality someone is aware of is an object of a unified awareness of both that quality and a color quality). This understanding obscures the distinction between awareness and unity. The way to avoid this obscuration is to understand *A* as the conjunction of truths about which entities (namely which qualities) are objects of awareness and as not having conjuncts about how the awareness of said objects is unified.

[12] For example, to avoid trivializing the entailment from *P* to *Q*, truths that mention the co-existence of physical truths alongside qualitative truths need to be excluded from figuring as conjuncts in *P*. Similarly, negative sentences such as "nothing instantiates a certain qualitative property" need to be excluded from *Q*—otherwise, the possibility of $P \sim Q$ may not entail the falsity of physicalism. See Chalmers (2010, p. 143) for two ways of dealing with the latter problem. Indexical truths, if they are to be regarded as distinct from physical truths but not of the sort that falsify physicalism, also require treatment—see Chalmers (2010, pp. 160–163) for suggestions for how to handle them.

sentence is (in this sense) conceivable iff there is an imaginable[13] scenario (1) which would, if actual, render the sentence true, and (2) no amount of ideal[14] reflection would reveal the scenario to be incoherent [see Chalmers (2010, pp. 143–144)]. It is worth noting, however, that there is a(n arguably) weaker form of conceivability that is available as a natural fallback: a sentence is *ideally negatively conceivable* iff it cannot be ruled out by ideal a priori reasoning (*ibid*). Used without qualification, 'conceivability' will henceforward mean ideal positive conceivability. I will attribute conceivability both to sentences and to scenarios, the latter sort of attributions being elliptical for the former.

P1 would be established by the conceivability of a conscious being in the actual world having a physical duplicate which differs by lacking at least some of the qualitative, awareness, unity, *or* possession properties of its conscious counterpart. I suspect that zombies are typically conceived of as lacking *all* such properties. Notice, however, that this premise only requires the conceivability of a world lacking *some* such properties. Since this premise is logically weaker than those playing its role in standard zombie arguments, it makes for a more powerful zombie argument that cannot be answered by simply defending the inconceivability of zombies that lack all such properties.

P2 claims that the conceivability of $P \sim (QAUS)$ implies its possibility. Premises in zombie arguments that link conceivability and possibility are typically supported by arguments for a general link between conceivability and possibility. The exact nature of this link is controversial. Chalmers frames his considered zombie argument in terms of two-dimensional semantics. Using that framework, he argues that a specific kind of conceivability (namely primary conceivability) entails a specific kind of possibility (namely 1-possibility, and, when the primary and secondary intensions of a sentence coincide in the right way, 2-possibility).

For tractability, I set aside sophisticated treatments of the conceivability-possibility link, and simply help myself to premises of the form: If $X$ is conceivable, then $X$ is possible. Opting for the simplified premises will ensure that the intuitive thrust of my arguments is conveyed, albeit without maximal precision. It should be borne in mind that the arguments presented in the main text admit of more sophisticated and defensible formulations.

P3 should be accepted on all sides: for if $P \sim (QAUS)$ is possible, then, contra physicalism, the fundamental physical truths about the actual world do not ground all truths. Thus, to reject the conclusion, one must reject P1 (as a priori/type-A physicalists do) or P2 (as a posteriori/type-B physicalists do).[15]

[13] One of the drawbacks of ideal *positive* conceivability is that the sense of imagination in terms of which it is defined is, even if intuitive, difficult to nail down precisely. [See (Goff (2009, p. 295, fn7) for more on this point, and Chalmers (2002) for an attempted characterization of the type of imagination in question.] I will not attempt to make progress on this issue here.

[14] A worry that is sometimes raised is that we are not in a position to evaluate the ideal conceivability of scenarios because we are not ideal agents. See Chalmers (2010, p. 155) for a reply.

[15] Or identify an ambiguity in the argument and claim that which premise is false depends on which disambiguation is adopted, as Russellian monists sometimes do.

As I show in the next section, this sort of reasoning can be extended to turn zombies against strict dualism.

### 3.3 Zombie arguments against strict dualism

I will present three zombie arguments against strict dualism. The three arguments are structurally similar, but they are partially independent. Roughly, the arguments work by iterating the conceivability–possibility reasoning employed within the zombie argument against physicalism twice over, once to metaphysically distinguish the physical from some non-physical type and a second time to distinguish both those types from another non-physical type, thereby establishing, contra strict dualism, the need for more than two general irreducible types to ground all truths.

Here is the first argument:

**D1**. $P \sim Q$ is conceivable.
**D2**. If $P \sim Q$ is conceivable, then $P \sim Q$ is possible.
**Lemma I**: So, $P \sim Q$ is possible.
**D3**. $PQ \sim A$ is conceivable
**D4**. If $PQ \sim A$ is conceivable, then $PQ \sim A$ is possible.
**Lemma II**: So, $PQ \sim A$ is possible.
**D5**. If $P \sim Q$ is possible and $PQ \sim A$ is possible, then strict dualism is false.
**C**. So, strict dualism is false.

The argument is valid: D1 and D2 imply Lemma I; D3 and D4 imply Lemma II (cf. Chalmers (2015, p. 273)); and, given D5, Lemmas I and II imply C. Let's go through the premises one at a time.

D1 asserts the conceivability of $P \sim Q$, and hence that there is a scenario (coherent on ideal reflection) in which all physical truths about the actual world obtain and yet in which not all qualitative truths about the actual world obtain. Admittedly, there is only so much one can do to evoke the recognition of conceivability in someone who does not recognize it without prompting. Nevertheless, there are two procedures worth mentioning that may induce recognition.

First, one starts a conception from scratch and tries to *construct* the target scenario by sequentially adding elements to the conception. For example, in this case one might first add to a blank-slate conception by conceiving of space–time without conceiving of it as having contents. Then, augment one's conception by imagining space–time to be populated with the other fundamental physical entities. Next, add the distribution of physical properties that obtain in the actual world.[16] If all goes well (as it did in my case, and hopefully yours too), the scenario thusly conceived is a coherent one in which not all qualitative truths about the actual world obtain.

[16] Of course, because one does not know all the physical truths about the actual world, one does not know the exact physical details that should be built into the scenario under consideration. The proponent of this argument will insist that this does not pose a serious objection since it is clear enough that however the physical details are filled in one will not be forced to bring the qualitative features of the actual world along with them.

Similarly, one could apply the constructive procedure to conceive of a zombie that is a physical duplicate of a conscious being in the actual world that falsifies $Q$. For example, suppose that a conscious being in the actual world instantiates qualitative blueness when it is having an experience as of a blue sky. To conceive of its zombie counterpart with the constructive procedure, one may sequentially conceive of the region it occupies, the physical entities in that region which compose the zombie-to-be, and their physical properties. One can stop there, noting that the creature thusly conceived lacks qualities, and hence falsifies $Q$. Alternatively, one could keep going, and add qualities to the zombie that differ from those instantiated by its conscious counterpart (and omit the ones had by the conscious counterpart)—such a zombie would also falsify $Q$. Appropriately plop this zombie into a scenario initially conceived of as a mere physical duplicate of the actual world, and one will have conceived of $P \sim Q$.

The second procedure starts with a conception of a scenario in which more truths obtain than in the target scenario, and then sequentially *abstracts* features from the initially conceived scenario until a conception of the target scenario is achieved. In the case at hand, one might start by imagining the scenario expressed by $PQ$, i.e. one in which all the physical and qualitative truths about the actual world obtain. One then excises at least some qualitative features of the initially conceived scenario. (N.B. one should conceive of these features as not existing, and not merely fail to conceive of them as existing.) For example, one might, holding the physical fixed, imagine a qualitatively crimson object in the actual world as being no color at all, or as being scarlet instead. Upon completing this procedure, one has imagined a scenario in which all physical truths, but not all qualitative truths, about the actual world obtain.

Given that conscious beings instantiate qualitative properties, one can apply the abstractive method to yield a zombie that would falsify $Q$. For example, start with a conception of one's current state, including the unified phenomenal state one is now enjoying. Attend to a small qualitative patch of one's visual field. Now, eliminate the patch from one's conception, while keeping the physical elements of the conception fixed. Plausibly, one is left with a coherent conception; and, if so, one has conceived of a physical duplicate that is not a qualitative duplicate.

So, given that the abstractive or constructive method succeeds, D1 is true, as $P \sim Q$ is conceivable. Next, D2 asserts that $P \sim Q$'s conceivability implies its possibility, i.e. that some scenario of the sort just conceived is possible. As noted above, I am taking premises of this form for granted. Note, however, that strict dualists need to accept D2 on pain of giving up zombie arguments against physicalism, the conceivability–possibility premises of which are justified on general grounds. D1 and D2 imply Lemma 1.

The next premise is D3, which claims that $PQ \sim A$ is conceivable, i.e. that there is a coherent scenario verifying all physical and qualitative truths and falsifying some truth about which qualities are objects of awareness. To imagine a scenario in which that is the case, we can again use the constructive method or the abstractive method. (For concision, I just use the former.)

As before, we proceed by sequentially constructing a scenario with space–time, the entities space–time contains, and their physical properties. The first time around,

we stopped building there and noted that the scenario thereby constructed did not contain the qualitative features of the actual scenario, and hence that we had conceived of a scenario at which $P \sim Q$ is true. In this case, the target scenario is one at which $PQ \sim A$ is true. So, we need to build the qualitative truths that obtain in the actual world into the scenario under construction.

Admittedly, because it is not obvious *where* (if anywhere) qualities are in the actual world, it is not obvious how we should distribute qualitative properties in the target scenario. However, examining candidate locations of qualities reveals that, regardless of whether qualities are located and if so where, we can conceive of them as existing absent awareness of them, and hence absent the awareness implied by $A$.

For example, suppose qualities are generally located where they often seem to be, namely on the surfaces of macroscopic objects. Could which qualities there is awareness of be derived a priori from the physical truths and the truths about the distribution of qualities over macroscopic objects? Such a derivation strikes me as impossible; indeed, the conception of such a scenario allows for a substantial degree of freedom in how awareness is distributed over it and it is implausible that this freedom would disappear entirely on ideal reflection. Analogous remarks apply if we instead suppose that qualities are in fact instantiated by the nervous systems of organisms, instead of the surfaces of objects. Which, if any, of these qualities are objects of awareness is up to the agent who, holding the physical and qualitative fixed, imagines the scenario.

Matters are less straightforward on the hypothesis that qualities (or at least the ones of which awareness is had) are not located in space–time. On that hypothesis, can we really form a conception in which awareness varies while the qualitative and physical is held fixed? In response, the answer is not obviously "yes" because it is not obvious that qualities can be conceived as existing outside of space–time; but, to the extent that I can imagine qualities as existing outside of space–time, I can also imagine the distribution of awareness varying while the qualitative and physical are held fixed. So, the noted hypothesis does not threaten the conceivability of $PQ \sim A$.

I know of no other plausible proposals about the distribution of qualities. Since $PQ \sim A$ is conceivable on any of these proposals and they are the only available plausible ones, I conclude that D3, which asserts the conceivability of $PQ \sim A$ should be accepted. The next premise is D4: If $PQ \sim A$ is conceivable, then $PQ \sim A$ is possible. As before, I am taking such premises—which antiphysicalists, per their use of general considerations to justify corresponding premises in zombie arguments against physicalism, are under pressure to accept—for granted. D3 and D4 imply Lemma II: $PQ \sim A$ is possible.

Notice that the just established Lemmas jointly imply the antecedent of D5: If $P \sim Q$ is possible and $PQ \sim A$ is possible, then strict dualism is false. Why accept D5? Well, strict dualism is committed to the fundamental truths of exactly two general types metaphysically entailing all truths about the actual world. But, to extend Kripke's (1980) metaphor, if $P \sim Q$ is possible, then God had to do more work after creating the world's fundamental physical nature. Specifically, he had to create the distribution of fundamental qualities. In that case, there are fundamental truths of at least two general types—namely the physical and qualitative—needed to

ground all truths.[17] But if $PQ \sim A$ is possible, then such truths are not metaphysically sufficient for all the truths. For God had to do even more work after he endowed the world with quality and physicality: he had to confer awareness upon it. Since fundamental awareness truths are ones of a third general type, if they (along with the physical and qualitative truths) are needed to metaphysically entail all truths, then strict dualism is mistaken in claiming that fundamental truths of just two general types metaphysically entail all truths. And since awareness is a third general type if $P \sim Q$ is possible and $PQ \sim A$ is possible, those possibilities jointly imply the falsity of strict dualism. That, of course, is exactly what D5 claims.[18]

That completes my defense of the first argument. Here is the second argument:

**Lemma II**: $PQ \sim A$ is possible.
**D6**. $PQA \sim U$ is conceivable.
**D7**. If $PQA \sim U$ is conceivable, then $PQA \sim U$ is possible.
**Lemma III**: So, $PQA \sim U$ is possible.
**D8**. If $PQ \sim A$ is possible and $PQA \sim U$ is possible, then strict dualism is false.
**C**. So, strict dualism is false.

Lemma II is taken from the previous argument. D6 asserts the conceivability of $PQA \sim U$, i.e. that there is, on ideal reflection, a coherently imaginable scenario that verifies all (actual) physical, qualitative, and awareness truths and falsifies some

[17] Strictly speaking, one could hold that the possibility of $P \sim Q$ does not imply that there are fundamental qualitative truths on the ground that the qualitative may follow from fundamental proto-qualitative truths that are neither physical nor qualitative. I suppress this type of qualification below, as proto-truths of type $X$ will generally pose just as much of a problem for strict dualism as fundamental truths of type $X$.

[18] Here is a two-dimensional semantics version of the argument, modeled after Chalmers's (2010, p. 149) two-dimensional argument against materialism:

**D1**. $P \sim Q$ is primarily conceivable.
**D2.0**. If $P \sim Q$ is primarily conceivable, then $P \sim Q$ is 1-possible.
**D2.5**. If $P \sim Q$ is 1-possible, then $P \sim Q$ is 2-possible.
**Lemma I***: So, $P \sim Q$ is 2-possible.
**D3**. $PQ \sim A$ is primarily conceivable.
**D4.0**. If $PQ \sim A$ is primarily conceivable, then $PQ \sim A$ is 1-possible.
**D4.5**. If $PQ \sim A$ is 1-possible, then $PQ \sim A$ is 2-possible.
**Lemma II***: So, $PQ \sim A$ is 2-possible.
**D5**. If $P \sim Q$ is 2-possible and $PQ \sim A$ is 2-possible, then strict dualism is false.
**C**. So, strict dualism is false.

Similar two-dimensional semantic formulations could be given for the other zombie arguments below. For concision, I omit such formulations. Chalmers's strategy for justifying premises such as D2.5 and D4.5 is to argue for the (at least partial) coincidence of sentences' primary and secondary intensions; however, as Chalmers (*ibid*, pp. 149–152) recognizes, such premises are susceptible to objections from Russellian monism, which reject the intensional coincidence of the sort Chalmers appeals to in order to justify those premises. The strict dualist may appropriate the Russellian monist objection to such premises, but she must do so selectively, lest she give up strict dualism for Russellian monism. Partly because not all Russellian monist style objections are equally plausible, it is not obvious whether the strict dualist can defend her position by selectively appropriating Russellian monist objections in a non-ad hoc fashion. I hope to discuss these matters in more detail on another occasion.

unity truth. So, we need to conceive of a zombie scenario that is a physical-qualitative-awareness duplicate that differs unity-wise.

To do that, we can start by conceiving of a merely physical-qualitative duplicate of the actual world. Next, we augment our conception by supposing that the qualities of which awareness is had in the actual world are objects of awareness in the scenario under construction. Now, whereas mere physical duplicates and mere physical-qualitative duplicates of the actual world are conceivable, a mere physical-qualitative-awareness duplicate is not (obviously) conceivable. The trouble is that it is unclear whether, and if so, to what extent, there is a requirement (accessible a priori) on awareness that it exist as (part of) a unity.

For example, could the awareness of an itch sensation conceivably exist all by itself, without belonging to some larger, unified state of awareness? The answer isn't obvious. To avoid having to take a stand on this issue, we should construct a scenario verifying $PQA \sim U$ by carefully redistributing unity relations over the scenario under construction, making sure that we do not leave outside of a unity relation any instance of awareness in the actual world whose ability to exist outside a unity relation is questionable. One way to do this is to conceive of the world being just as it is, except that your current unified state of awareness is replaced by two, perhaps one consisting of (part of) your actual visual experience and the other consisting of the rest of your total unified phenomenal state. (There are indefinitely many examples that could be used here; substitute your own if you don't like mine.)[19]

These considerations support D6. D7 ("If $PQA \sim U$ is conceivable, then $PQA \sim U$ is possible") is another premise linking conceivability to possibility. As noted, I take these premises—which antiphysicalists are under strong pressure to accept anyway—for granted. D6 and D7 imply Lemma III: $PQA \sim U$ is possible.

The next premise is D8, which claims that if $PQ \sim A$ is possible and $PQA \sim U$ is possible, then strict dualism is false. The rationale for D8 parallels that of D5 in the previous argument: strict dualism is committed to the fundamental truths of exactly two general types metaphysically entailing all truths about the actual world. But, (extending Kripke's metaphor again) if $PQ \sim A$ is possible, then God had to do more work after creating the world's fundamental physical-qualitative nature. Specifically, he had to create the distribution of awareness. In that case, there are fundamental truths of at least two general types of truths—namely the physical-qualitative type and the awareness type—needed to ground all truths. But if $PQA \sim U$ is possible, then such truths are not metaphysically sufficient for all the truths. For God had to do even more work after he endowed the world with physicality, quality, and awareness: he had to impose the unity relation upon it.[20] Since fundamental unity truths are ones of a third[21] general type, if they (along with

[19] Also, remember that the argument can be run in terms of ideal negative conceivability if ideal positive conceivability fails.

[20] Well, if awareness requires unity then God couldn't have *created* awareness without also creating unity. Still, if he needn't have created any specific unity relation over the awareness in the world, then after deciding which qualities were to enjoy the privilege of being objects of awareness, he had to make a further *decision* about which unity relation to confer upon it.

[21] Or fourth, if the physical and qualitative are also distinct types.

the physical-qualitative truths and the awareness truths) are needed to metaphysically entail all truths, then strict dualism is mistaken in claiming that fundamental truths of just two general types metaphysically ensure all truths. And since unity is a third general type if $PQ \sim A$ is possible and $PQA \sim U$ is possible, those possibilities jointly imply the falsity of strict dualism. That, of course, is exactly what D8 claims.

That concludes my presentation of the second zombie argument against strict dualism. Here is the third:

**Lemma III**: $PQA \sim U$ is possible.
**D9**. $PQAU \sim S$ is conceivable.
**D10**. If $PQAU \sim S$ is conceivable, then $PQAU \sim S$ is possible.
**Lemma IV**: So, $PQAU \sim S$ is possible.
**D11**. If $PQA \sim U$ is possible and $PQAU \sim S$ is possible, then strict dualism is false.
**C**. So, strict dualism is false.

Lemma III is taken from the previous argument. D9 asserts the conceivability of $PQAU \sim S$, i.e. that there is, on ideal reflection, a coherently imaginable scenario that verifies all (actual) physical, qualitative, awareness, and unity truths and falsifies some possession truth. So, we need to construct a zombie scenario that is a physical-qualitative-awareness-unity duplicate that differs possession-wise. Recall that possession consists in qualities, awareness of qualities, or unified awareness of qualities being possessed by an entity (type).

Here is one way to go about conceiving of the target scenario: start with a full conception of the actual world, including its physical, qualitative, awareness, unity, and possession features. Next, while holding its other features fixed, modify the distribution of possession relations over the scenario so as to falsify $S$. There are a number of modifications that seem to do the trick. One consists in the complete elimination of possession relations from the scenario.[22]

Less ambitiously, one might modify the conception of the actual scenario into a conception of the target scenario by eliminating some possession relations while keeping others fixed. For example, if a color quality is actually possessed by an object, then one can modify one's conception by conceiving of the quality as co-located with the object, but not possessed by it.

Still less ambitiously, one might replace some possession relations, while keeping others fixed. For example, one could modify the scenario so that a quality that is in fact instantiated by an object but not a region is instantiated by the region in which the object is located, but not the object. Similarly, one might conceive of

[22] The actually possessed qualities, awareness, and unity would then exist unpossessed in the target scenario. Some might find this objectionable because it conflicts with the dictum "no property instantiation without a bearer". I see no reason to accept this dictum. (A merely apparent reason is its (superficial) similarity with the dictum "No relations without relata"—*that* dictum is quite plausible. However, eliminativism about relata but not relations is incoherent, whereas eliminativism about bearers but not property instantiations is not.) In any case, subsequent examples will verify $PQAU \sim S$ without conflicting with the noted dictum.

one's current unified conscious state as being possessed by space–time, a fundamental particle, or a slightly expanded or contracted physical base as opposed to whatever actually instantiates it.

These examples exhibit zombies (in my broad sense of the term) of the sort that falsify $S$ (in the scenarios they inhabit), and, if appropriately embedded within scenarios that are physical-qualitative-awareness-unity duplicates of the actual world, verify $PQAU \sim S$. It is worth emphasizing that D9 only requires the conceivability of one zombie scenario that verifies $PQAU \sim S$. The previous paragraph describes a number of such scenarios, and suggests indefinitely many. To deny D9, one must (implausibly) deny that any such scenario is coherently imaginable. (And even that would not be enough to put the argument to rest. For D9 could be re-construed in terms of ideal negative conceivability, in which case denying D9 would commit one to all such scenarios being ruled out a priori on ideal reflection.) For these reasons, D9 should be accepted.

D10 is yet another premise linking conceivability and possibility. As one might have guessed by now, as a premise of that sort, I will assume D10 without argument. D9 and D10 jointly imply Lemma IV: $PQAU \sim S$ is possible.

The final premise is D11, which claims that if $PQA \sim U$ is possible and $PQAU \sim S$ is possible, then strict dualism is false. The rationale for D11 parallels that of D5 and D8 in the previous two arguments: strict dualism is committed to the fundamental truths of exactly two general types metaphysically entailing all truths about the actual world. But, if $PQA \sim U$ is possible, then God had to do more work after creating the world's fundamental physical-qualitative-awareness nature. Specifically, he had to generate a fundamental unity relation over the awareness features of the world. In that case, there are fundamental truths of at least two general types of truths—namely the physical-qualitative-awareness type and the unity type—needed to ground all truths. But if $PQAU \sim S$ is possible, then such truths are not metaphysically sufficient for all the truths. For God had to do even more work after he decided to endow it with physicality, quality, awareness, and unity:[23] he had to decide which (types of) entities would possess which of these features. Since fundamental possession truths are ones of another general type, if they (along with the physical-qualitative-awareness truths and the unity truths) are needed to metaphysically entail all truths, then strict dualism is mistaken in claiming that fundamental truths of just two types ground all truths. And since possession is another general type if $PQA \sim U$ is possible and $PQAU \sim S$ is possible, those possibilities jointly imply the falsity of strict dualism. That, of course, is exactly what D11 claims.

The individual and collective upshot of these three arguments is bad news for strict dualists: if strict dualists wield a zombie argument against physicalism, they open themselves to zombie attacks on (at least) three fronts. On their own, the

[23] If some truthmakers of *PQAU* couldn't exist without possession relations, then God couldn't have *created* those truthmakers without creating possession relations. Still, if he needn't have created any specific set of possession relations, then after deciding how to distribute the truthmakers of *PQAU*, he had to make a further *decision* about which possession relations to impose upon it.

zombie arguments presented in this section make a strong, direct case against strict dualism. But these arguments also indirectly threaten strict dualism once the dialectical context is taken into consideration. For given the similarities between the zombie argument against physicalism and the zombie arguments against strict dualism, it seems safe to assume that a successful rebuttal of all zombie arguments against strict dualism would bring with it the resources for rebutting the zombie argument against physicalism. Since the zombie argument against physicalism is a chief motivation for antiphysicalism, giving up on zombie arguments effectively requires giving up on antiphysicalism, and hence strict dualism.[24] So, strict dualists are neither in a position to accept nor reject zombie arguments: acceptance is straightforwardly incompatible with their position and rejection leaves their position unmotivated. I conclude, in light of these difficulties, that strict dualism should be rejected.

## 4 Can strict dualists forswear zombies while retaining motivation?

I suggested above that the strict dualists (in particular and antiphysicalists more generally) cannot simply walk away from zombie arguments, as they need such arguments to motivate their position. It might be objected that such theorists do not need zombie arguments to motivate their position: they can use non-zombie arguments against physicalism to motivate their position instead.

While I believe that this objection deserves serious consideration and development, I am pessimistic about its prospects for two reasons. First, it is not obvious that the arguments against physicalism that have received significant discussion in the literature—e.g., the explanatory gap argument, Jackson's (1982) knowledge argument, Kripke's (1980) modal argument, White's (2007) property dualism argument, and Bealer's (2002) semantic stability argument—advance considerations that are independent of those advanced by the zombie argument. For, as Chalmers (2010, pp. 192–205) has shown, these arguments can be given a unified treatment within the framework of two-dimensional semantics. It would be very surprising if these arguments against physicalism failed to advance independent considerations but were nonetheless such that only some of them had analogs that threatened strict dualism. (Examining these arguments individually to determine whether they have analogs exceeds this paper's scope.)

Second, even if other arguments against physicalism are independent of the zombie argument, there are good prima facie reasons for thinking that such arguments have analogs that threaten strict dualism. For example, any explanatory gap argument against physicalism would have to assume that the inexplicability of

[24] In contrast, since zombie arguments against strict dualism are not a chief motivation of physicalism, the physicalist does not risk undermining the motivation for her position by rejecting such arguments as unsound. If she does so, however, she cannot wield such arguments against strict dualists. Still, she can do something almost as good for her position: she can maintain that the strict dualist must treat such arguments as direct threats if they endorse a zombie argument against physicalism, albeit not as direct threats to strict dualism issued by physicalists.

one type in terms of another implies that the latter does not ground the former. But, as we saw in Sect. 2, there are a number of explanatory gaps between (combinations of) the physical, qualitative, awareness, unity, and possession truths. So, if explanatory gaps demarcate metaphysical distinctions at all, they demarcate more than the strict dualist can countenance.

Similarly, consider Jackson's (1982, p. 130) Mary, who knows all the physical truths, but who (putatively) gains new knowledge (e.g. about what it is like to see ripe tomatoes) when she is released from her black and white room.[25] If Mary's new knowledge is a problem for physicalism, then there are analogs of Mary whose new knowledge is a problem for strict dualism. I will describe a couple of such analogs.

First, consider a scenario inhabited by Mary's analog, $\text{Mary}_{\text{P}}$: like Mary, $\text{Mary}_{\text{P}}$ is a brilliant scientist (indeed, she is an ideal reasoner) who has lived her whole life in the confines of a black and white room. During her time in the room, she has come to know all the physical truths (and their a priori consequences). Moreover, during her time in the room she built (never mind how) a pair of qualitative-vision goggles that enables her to have accurate visual experiences of qualities. What will happen when she leaves the room and puts on the goggles? Just as it is intuitively obvious that Mary learns something when she leaves the room for the first time, so too is it obvious $\text{Mary}_{\text{P}}$ learns something about the qualitative nature of the outside world when she leaves.

Second, Consider $\text{Mary}_{\text{PQ}}$: she too is an ideal reasoner who has lived her whole life in a single room. She has come to know all the physical and qualitative truths about a possible world *w*. She has learned the qualitative truths via a color computer screen in her room. As it turns out, she has noticed that *w* contains a room just like hers in qualitative and physical respects down to the last detail. Naturally, it contains a physical-qualitative duplicate of her.

Having learned all the physical and qualitative truths about *w*, one day she becomes bored. To cope with her boredom, she takes up philosophy. She starts by asking whether any creatures in *w* are aware of anything. Assuming, as would intuitively seem to be the case, she cannot figure out the answer to this question by reasoning from the information she already had, we may suppose that she emails the scientific organization that has provided her with all the physical and qualitative truths about *w*. In the email, she asks for information about whether any of the creatures in *w* exhibit awareness. The organization (truthfully) responds as follows: "you live in *w*."

What happens when $\text{Mary}_{\text{PQ}}$ reads this email? Intuitively, upon reading the email, $\text{Mary}_{\text{PQ}}$ at least learns that one of *w*'s creatures enjoys awareness.[26] I submit that the intuitions that Mary, $\text{Mary}_{\text{P}}$, and $\text{Mary}_{\text{PQ}}$ acquire new knowledge (of what it's like, of quality, and of awareness, respectively) are robust across these cases. Moreover, given the similarity between the cases, if the Mary scenario shows that

[25] It's worth noting that Jackson (2003) no longer accepts the knowledge argument.

[26] Plausibly, she gains new indexical knowledge as well (e.g. of the truth that *she* is in *w*). However, equally plausibly, this is not the only kind of knowledge she gains: the truth that someone is aware in *w* is not an indexical truth, but $\text{Mary}_{\text{PQ}}$ comes to know it. Thus, this knowledge argument is not defused by the "$\text{Mary}_{\text{PQ}}$ merely gains *de se* knowledge" response.

the physical truths do not ground all truths, then the $\text{Mary}_{\text{P}}$ scenario shows that the physical truths do not ground the qualitative truths and the $\text{Mary}_{\text{PQ}}$ scenario shows that the physical and qualitative truths do not ground the awareness truths. But if the physical truths do not ground the qualitative truths and the physical and qualitative truths do not ground the awareness truths, then strict dualism is false. So, on pain of making her position susceptible to a knowledge argument, the strict dualist cannot use the knowledge argument against physicalism to motivate her position.

To sum up, pending argument to the contrary, I am skeptical that strict dualism can be motivated by non-zombie arguments against physicalism in the literature both because these arguments are not obviously independent of zombie arguments and because even if they are, it is not difficult to construct analogs of them that land strict dualists in the same dialectical predicament as zombie arguments.

## 5 The prospects for pluralism

At this point, the antiphysicalist might be tempted to abandon strict dualism for pluralism, which holds that the fundamental truths of more than two general irreducible types ground all truths. In this section, I develop an argument against pluralism.

Recall the core argument:

1. If strict dualism and pluralism should be rejected, antiphysicalism should be rejected (and physicalism accepted).
2. Strict dualism should be rejected.
3. Pluralism should be rejected.
4. So, antiphysicalism should be rejected (and physicalism accepted).

So far, I have shown how premises 1 and 2 can be defended. After supporting premise 3 in this section, my presentation of the core argument will be complete and I will have exhibited a new route to 4.

My proposed defense of 3 proceeds in two steps. First, I argue that, to retain a motivated position, pluralists must adopt *radical pluralism,* which holds that the fundamental truths of significantly more than two general irreducible types are needed to ground all truths. Second, I will consider radical pluralism's expansive ontology as a basis for its rejection.

### 5.1 Why retreating to pluralism requires radicalization

Pluralism's ontology is more expansive than strict dualism's, but it need not be *much* more expansive. If a bit of ontological inflation could cure the antiphysicalist of the strict dualist's ills, pluralism might seem to be just what the doctor ordered. In this section, I argue that no such moderate pluralist escape is available to the antiphysicalist: while the ills of strict dualism can be cured with ontological inflation, they cannot be cured with a small dose. The only cure comes in the form of radical pluralism, which has the unfortunate side-effect of substantial ontological inflation.

This conclusion can be derived by examining the below argument, which uses premises from the arguments in Sect. 3:

**Conceivability–possibility principle (CPP)**: If $X$ is conceivable, $X$ is possible.
**D1**. $P \sim Q$ is conceivable.
**Lemma I**: So, $P \sim Q$ is possible. [*D1 and CPP*]
**D3**. $PQ \sim A$ is conceivable.
**Lemma II**: So, $PQ \sim A$ is possible. [*D3 and CPP*]
**D6**. $PQA \sim U$ is conceivable.
**Lemma III**: So, $PQA \sim U$ is possible. [*D6 and CPP*]
**D9**. $PQAU \sim S$ is conceivable.
**Lemma IV**: So, $PQAU \sim S$ is possible. [*D9 and CPP*]
**Radicalization**: Radical pluralism is true if each of the following is possible: $P \sim Q$, $PQ \sim A$, $PQA \sim U$, and $PQAU \sim S$.
**Conclusion**: So, radical pluralism is true. [*Lemmas I–IV and Radicalization*]

Since this argument is valid, the pluralist who wishes to avoid radicalization must deny one or more of this argument's premises. What are the pluralist's options?

First, she might reject CPP.[27] This option is a non-starter. For the hope in seeking a pluralist alternative to strict dualism was that zombies might be retained in the service of antiphysicalism. But to reject CPP is to forgo the means by which zombies are employed, and hence to forgo zombies—if the antiphysicalist were going to make do without zombies, she would be better off sticking with strict dualism, as she would get nothing in return for purchasing the pluralist's inflated ontology.

Second, the pluralist might reject one or more of the conceivability premises. But Sect. 3 offered compelling reasons for accepting each of these premises. Taking this option would require the pluralist to explain why those reasons are misguided.

Third, the pluralist might reject Radicalization. This might seem like the most promising option since I have not said anything to support Radicalization yet. In fact, however, I do not believe there is any way to plausibly deny Radicalization on my intended understanding of it. Here's why: suppose that $P \sim Q$, $PQ \sim A$, $PQA \sim U$,

[27] She might try to do this by appropriating the phenomenal concept strategy for defending physicalism. When directed at zombie arguments against physicalism, this strategy (of which there are number of versions) proceeds by trying to identify a distinctive feature of phenomenal concepts that explains why zombies would be conceivable even if physicalism were true, thereby undermining the inference from the conceivability of zombies to their possibility. [E.g. see Balog (2012).] Similarly, the pluralist might try to resist the applications of CPP within the above argument by appealing to distinctive features of our qualitative, awareness, unity, and possession concepts to explain why the scenarios invoked by the argument would be conceivable even if they were impossible. Though this response deserves further investigation, I am pessimistic about it for two reasons. First, the pluralist would have to apply this strategy selectively, lest her view collapse into a physicalist version of the phenomenal concept strategy. But I do not see how the strategy could be applied selectively in a non-ad hoc fashion. Second, the phenomenal concept strategy is motivated by the thought that there is something *distinctive* about phenomenal concepts that generates the explanatory gap between the phenomenal and the physical. This motivation is undermined by the observation that there are in fact many such gaps in the vicinity (see Sect. 2) and that these are generated by (at least conceptually) diverse phenomena (namely, the physical, qualitative, awareness, unity, and possession truths).

and $PQAU \sim S$ are each possible. In that case, to once again use Kripke's metaphor, God had to do a lot of work to create our world. First, he had to give it physicality. Since, $P \sim Q$ is possible, giving the world physicality was not enough to give it qualities; so God had to engage in a second act of creation by distributing qualities. But, since $PQ \sim A$ is possible, he could not stop there: he had to take a third step, deciding which qualities were to be objects of awareness. And since $PQA \sim U$ is possible, three steps were not enough: a fourth was needed to fix the unity truths; and finally, a fifth was needed to fix the possession truths. Corresponding to the five steps are (given the noted possibilities) five general irreducible types, the fundamental instances of which are needed to fix all truths. Five is—I will assume in this context—significantly more than two. So, as Radicalization claims, given the possibility of each of $P \sim Q$, $PQ \sim A$, $PQA \sim U$, and $PQAU \sim S$, radical pluralism is true, as there are significantly more than two general irreducible types of truth whose fundamental instances are needed to ground all truths.

Since the pluralist is not in a position to resist the argument by exercising any of the available options, I conclude that pluralists should be radical pluralists.

### 5.2 The cost of radicalization

While I am not ordinarily gripped by the minimalist, ontological splendor of desert landscapes, I must admit that even I find the radical pluralist's ontological exuberance distasteful—five general irreducible types seems like an extravagant posit that calls out for ontological consolidation. Yet, no such consolidation is possible if, as I am inclined to believe, a conceivability–possibility principle is true and my defense of the conceivability premises succeeds.

This is a welcome result for the physicalist: I do not think it would be irrational for her to reject the zombie argument against her position on the ground that radical pluralism and its inflated ontology are absurd and inevitable consequences of following that form of argument where it leads.[28] If one goes along with the physicalist here and takes radical pluralism's offense to parsimony as a sufficient reason for rejecting it, then it is a short step to reject pluralism more generally—one need only accept my argument in Sect. 5.1 that pluralists should be radical.

As someone who does not accept physicalism, this is my preferred place to get off the boat: though I admit its inflated ontology is a significant cost, I would be

[28] It is also open to the physicalist to supplement the parsimony objection to radical pluralism with a causal argument that replaces premises in the standard causal argument for physicalism [e.g. see Papineau (2002)] with weaker premises. Rather than arguing that the causal closure of the physical leaves no work for the non-physical to do (non-redundantly), the physicalist could argue that the causal closure of (say) the domain encompassing the physical, qualities, and awareness leaves no work for distinct phenomenal unity and possession relations to do. Whether such arguments would exert significantly more dialectical force against the radical pluralist than the standard causal argument depends on whether their premises are significantly more plausible—of course, this is not settled by the mere fact that such arguments contain weaker premises than the standard argument. I will not attempt to adjudicate this issue here.

willing to buy into radical pluralism if that is the price of being an antiphysicalist with a motivated position.[29]

## 6 Conclusion

To sum up, I have argued that strict dualism should be rejected because it is threatened by three zombie arguments and it cannot plausibly resist these without giving up the zombie argument against physicalism, thereby undermining its own motivation. I then showed how to make a case against pluralism: one first argues that, again on pain of giving up the zombie argument against physicalism and hence motivation for antiphysicalism, abandoning strict dualism for pluralism requires radicalization—one must adopt a pluralism that posits significantly more than two general types to ground all truths—which in turn leads to an (arguably) unacceptably inflated ontology.

If this argument succeeds in providing sufficient grounds for rejecting strict dualism and pluralism, then it provides sufficient grounds for rejecting antiphysicalism, as strict dualism and pluralism form a partition of antiphysicalist positions. And given that the fundamental physical truths are among those needed to ground all truths, the argument establishes physicalism if it rules out antiphysicalism.

At first glance, physicalists and zombies seem to be strange bedfellows. The physicalist must, after all, deny the possibility of zombies. But, as the argument I have developed in this paper shows, despite this adversarial aspect of their relationship, there is room for a physicalist-zombie alliance. While the physicalist needs to prevent zombies from entering modal space, it is to her advantage to invite zombies into dialectical space to feature in arguments against her antiphysicalist opponents.

**Acknowledgments** I am grateful to Jaegwon Kim for helpful comments on two predecessors of this paper, to Josh Dever, Nina Emery, Chris Hill, and Josh Schechter for helpful comments on one predecessor of this paper, and to Robert Joynt, Han Li, Adam Pautz, and Nick Smyth for helpful discussion.

[29] There is a temptation to downplay the size of radical pluralism's ontology by claiming that all or some of the types—quality, awareness, unity, and possession—to which radical pluralism is committed are species of a general type—e.g. the mental—to which all reasonable ontologies are committed, and that ontology expansion by way of mere species augmentation is a moderate and unobjectionable form of inflation. This temptation should be resisted: on pain of classifying all forms of ontological expansion as unobjectionable, a distinction needs to be drawn between species of unitary, natural (in the Lewisian sense) general types and species of general types that are not. Plausibly, admitting additional species of a unitary type to which we are already committed is a moderate form of ontology expansion. However, there does not seem to be such a unitary type of which qualities, awareness, unity, and possession are species. Of course, we can use or invent a general label (e.g. 'mental') that applies to all of them, but that alone neither reveals a unitary type to which the noted species belong nor absolves radical pluralism of severe parsimony violation.

## References


Balog, K. (2012). In defense of the phenomenal concept strategy. *Philosophy and Phenomenological Research, 84*(1), 1–23.

Bayne, T., & Chalmers, D. (2003). What is the unity of consciousness? In A. Cleeremans (Ed.), *The unity of consciousness*. New York: OUP.

Bealer, G. (2002). Modal epistemology and the rationalist renaissance. In T. Gendler & J. Hawthorne (Eds.), *Conceivability and possibility*. New York: OUP.

Bennett, K. (2011). By our bootstraps. *Philosophical Perspectives, 25*(1), 27–41.

Chalmers, D. (2002). Does conceivability entail possibility? In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and possibility* (pp. 145–200). New York: OUP.

Chalmers, D. (2010). *The character of consciousness*. New York: OUP.

Chalmers, D. (2015). Panpsychism and panprotopsychism. In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the physical world: Perspectives on Russellian monism* (pp. 246–276). OUP.

Chalmers, D. (forthcoming). The combination problem for panpsychism. In G. Bruntrup & L. Jaskolla (Eds.), *Panpsychism*. Oxford University Press.

Coleman, S. (2012). Mental chemistry: Combination for panpsychists. *Dialectica, 66*(1), 137–166.

Dasgupta, S. (2014). The possibility of physicalism. *Journal of Philosophy*, *111*(9/10), 557–592.

Goff, P. (2009). Why panpsychism doesn't help us explain consciousness. *Dialectica, 63*(3), 289–311.

Goff, P. (2012). Does Mary know I experience plus rather than quus? A new hard problem. *Philosophical Studies, 160*(2), 223–235.

Goff, P. (ms). *Chapter 2—What is physicalism?* http://www.philipgoffphilosophy.com/uploads/1/4/4/4/14443634/chapter_2_-_what_is_physicalism.pdf.

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

Jackson, F. (2003). Mind and illusion. In A. O'Hear (Ed.), *Minds and persons*. Cambridge: Cambridge University Press.

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

Lee, G. (2013). Materialism and the epistemic significance of consciousness. In U. Kriegel (Ed.), *Current controversies in philosophy of mind*. London: Routledge.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly, 64*, 354–361.

Lewis, D. (2009). Ramseyan humility. In D. Braddon-Mitchell & R. Nola (Eds.), *Conceptual analysis and philosophical naturalism* (pp. 203–222). Cambridge: MIT Press.

Montero, B. (2006). Physicalism in an infinitely decomposable world. *Erkenntnis, 64*(2), 177–191.

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

Nagel, T. (1986). *The view from nowhere*. New York: OUP.

Papineau, D. (2002). *Thinking about consciousness*. New York: OUP.

Pautz, A. (2009). A simple view of consciousness. In R. C. Koons & G. Bealer (Eds.), *The waning of materialism: New essays* (pp. 25–66). New York: OUP.

Sosa, D. (2011). Two forms of dualism. *Dialogue, 50*(02), 307–313.

White, S. (2007). Property dualism, phenomenal concepts and the semantic premise. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism*. New York: OUP.

Yetter-Chappell, H., & Chappell, R. Y. (2013). Mind-body meets metaethics: A moral concept strategy. *Philosophical Studies, 165*(3), 865–878.