Zombies and Epiphenomenalism

ANDREW BAILEY *University of Guelph*

ABSTRACT: This paper examines the relationship between the claim that zombies are logically/metaphysically possible and the position that phenomenal consciousness is epiphenomenal. It is often taken that the former entails the latter, and that therefore any implausibility in the notion of conscious epiphenomenalism calls into question the genuine possibility of zombies. Four ways in which the zombist might respond are examined, and I argue that two—those most commonly encountered—are inadequate, but the others—one of which is rarely formulated and the other new—are more forceful. The upshot, nevertheless, is that the zombist may indeed face an unwelcome commitment to conscious epiphenomenalism.

RÉSUMÉ : Cet article porte sur la relation entre l'assertion que les morts-vivants sont une possibilité logique/métaphysique et la position que la conscience des phénomènes est épiphénoménale. On suppose souvent que ceci implique cela, et que, par conséquent, tout caractère non plausible dans la notion de conscience épiphénoménale remet en question la possibilité que des morts-vivants existent vraiment. Suit l'examen de quatre voies ouvertes au zombiste pour formuler une réponse, deux desquelles sont déclarée inadéquates, tandis que les deux autres se révèlent avoir plus de force. Le résultat en est cependant que le zombiste se trouve malgré lui engagé à défendre la conscience épiphénoménale.

One of the chief munitions in the anti-physicalist armament is the zombie. Zombies are philosophical fantasy creatures that are physically just as we are (or at least, just as our best physicalist guess says we are) but that lack phenomenal consciousness. The zombist's claim is that the existence of zombies is a logical and possibly a metaphysical (though not an empirical) possibility; if zombies are logically possible, then it putatively follows that facts about (the presence or absence of) phenomenal consciousness are not fixed by the totality

***Dialogue* 48 (2009), 129-144. Printed in the U.S.A.**
© 2009 Canadian Philosophical Association/Association canadienne de philosophie
doi:10.1017/S0012217309090076

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

of physical facts, and thus that consciousness is *over and above* the physical; hence, physicalism is false.

Compare the cases of being a live pig, being liquid, or even, probably, being Michelangelo's *David*. In these cases, and all others relevantly like them, if the totality of physical facts are fixed then so too are the facts about life, pigs, liquidity, and the provenance and other (non–consciousness-dependent) qualities of artworks. A "zombie" world that is physically just like this one but in which nothing is alive—containing only the walking (breathing, reproducing, metabolising, etc.) dead, as it were—is not merely unlikely but internally incoherent. The alleged possibility of philosophical zombies is thus thought to identify something *special* about phenomenal consciousness, and to raise a compelling problem for physicalism.

The zombist case is far from won, however, and various physicalist responses to the zombie argument have been raised, particularly since the resurgence of interest in zombie-style arguments stimulated by David Chalmers' *The Conscious Mind* in 1996.[1] One particularly prevalent style of response on the part of the physicalist is to connect the logical possibility of zombies with the truth of *epiphenomenalism* about phenomenal consciousness. Insofar as epiphenomenalism is implausible, unpalatable, or even impossible, and insofar as the zombist claim leads to epiphenomenalism, thus far are zombie arguments rendered doubtful.[2] In its outlines, the physicalist response looks like this:

i) A pivotal premise in zombie arguments is that zombies are (in the relevant sense) possible.
ii) If zombies were possible, then phenomenal consciousness would be epiphenomenal.
iii) But phenomenal consciousness cannot be, or at least is not, epiphenomenal.
iv) Therefore zombies are not possible, and so the zombist challenge to physicalism fails.

Premise iii) can be, and has been, motivated in various ways. What all these have in common is the intuition that it is implausible or impossible to suppose that the elimination of phenomenal consciousness from the world would make no difference—that everything (including cognition) would go on just the same even if pain were not phenomenally painful, vision not phenomenally colourful, desire not phenomenally pressing, and so on. A particularly compelling instance of this general worry is what is sometimes called the Paradox of Phenomenal Judgement (after Chalmers 1996): the puzzle that, since cognition *ex hypothesi* remains unaffected by the removal of phenomenal consciousness, our first-personal, introspective beliefs about and reports of our own states of phenomenal consciousness would—if consciousness were epiphenomenal—prima facie be completely unchanged even in the absence of their objects. That

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

is, if consciousness is epiphenomenal, we would not and could not notice if *we ourselves* were zombies.

Although there is much more to be said on this score, I shall not concern myself here about the implausibility of epiphenomenal consciousness. This paper deals instead with premise ii) of the argument above.[3] What exactly is the connection between the zombie hypothesis and the epiphenomenalism of phenomenal consciousness? *Does* the mere logical or metaphysical possibility of zombies entail conscious epiphenomenalism? And if conscious epiphenomenalism is entailed, is what would follow the truth of epiphenomenalism in the actual world or merely the physical possibility of epiphenomenalism (i.e., its obtaining in at least one possible world physically indiscernible from the way the physicalist takes the actual world to be)? That is, must the physicalist show, to complete her argument, that conscious epiphenomenalism is *impossible* or merely that it is not the case? These issues, though crucial to the dialectic about zombies and phenomenal consciousness, have not been given much attention in the literature to date.

I shall argue that the relationship between the zombie hypothesis and epiphenomenalism is significantly more complicated than is usually recognised, but that when these complications are worked through the connection emerges as tight enough to be worrying for the zombist. Though not a particularly one-sided or clear-cut conclusion—and thus in *that* sense less interesting than it might be—I think it has the modest virtue of being plausible, and as nuanced as the multifaceted situation requires.

1. The Initial Intuition

It is often taken for granted that, modulo the odd unimportant complication, the logical/metaphysical possibility of zombies does indeed necessitate the epiphenomenalism of phenomenal consciousness.[4] Indeed, many proponents of the zombie argument—including at one time the ur-zombist David Chalmers—seem willing to accept this claim, at least provisionally, and devote their efforts to attempting to show that conscious epiphenomenalism is a bullet we should be willing to bite.[5]

It is natural to think of zombies in the following way: zombies are what are left over when you subtract phenomenal consciousness from the actual world, or a relevant part of it, while leaving everything physical unchanged. To say that zombies are logically/metaphysically possible is to say that it is logically/metaphysically possible to perform this subtraction. But to allow this is precisely to allow that the presence of phenomenal consciousness is unnecessary for any physically specifiable event to occur, since *ex hypothesi* all these events (or states, or properties) would occur anyway even if consciousness were not present, and furthermore that the occurrence of any or all of these physical events is not causally sufficient for the presence of consciousness, since they may occur when conscious events do not. That is, consciousness is strongly epiphenomenal—it neither causes nor is caused by any physical event.

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

Furthermore, according to this line of thought, consciousness is *actually* epiphenomenal, not merely epiphenomenal in some other possible world. Bear in mind that, for the zombie hypothesis to be meaningful as a problem case for physicalism with respect to consciousness, the zombist must allow the physicalist to specify physicalism on her own terms (rather than foist upon her a specification of physicalism that she may then reject). Thus, the zombie world is to include everything that the physicalist takes to be required for physicalism to be true, including not only the distribution and properties of all the physical entities but also all the physical laws that govern the interactions of those entities. The zombist contention is that, even *given* everything the physicalist might wish to include in a final completed physics, it is (logically or metaphysically) possible to subtract consciousness. So a zombie world cannot legitimately be created by adjusting the causal laws (at least if we follow the standard line in taking causation—though not necessarily the events related by causal laws—to be within the domain of physics). If physical events do not cause conscious events (or vice versa) in the zombie world, then they do not do so in the physicalist's actual world either.[6]

Bear in mind also that some sort of reference to the *actual* world in the specification of the zombie hypothesis is non-negotiable, since it is claims about consciousness in the actual world that are under dispute. It is uncontroversial that both 'zombies' and epiphenomenal consciousness are possible, in the sense that there are *some* possible worlds in which human-like organisms are not phenomenally conscious (perhaps because their heads are filled with sawdust) or consciousness is causally inert (perhaps because it is possessed only by disembodied spirits in that world). What is at issue is whether zombies are possible *given* what we know, or think we know, about the physics of the actual world—i.e., in any possible world "physically indiscernible" from the actual one.

2. Two Unpersuasive Responses

It is widely conceded by physicalists that the necessitation of conscious epiphenomenalism by the logical possibility of zombies falls short of perfect entailment. However, these initial stumbling blocks do not do much to ameliorate the power of the physicalist response to the zombist.

The first caveat is that the zombie possibility leads to epiphenomenalism only if causal overdetermination is not the case. Suppose a) that zombies are possible and hence that phenomenal consciousness is non-physical, and b) that certain events have both a sufficient physical cause and an additional phenomenal one. For example, suppose that particular brain states are alone sufficient to cause pain behaviour, but that this behaviour is *also* caused by (non-physical) phenomenal pain sensations. Under such circumstances, of course, consciousness would not be required for everything to happen in the actual world just as it does—i.e., zombies are possible—but nevertheless phenomenal consciousness would not be epiphenomenal.

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

But this does little or nothing to weaken the physicalist response. First, this zombist reply depends upon overdetermination being the case—since consciousness is non-epiphenomenal only if overdetermination is true for every instance of phenomenal causation—and this is not itself a very plausible position. More fundamentally, this complication is anyway irrelevant to the force of the physicalist response. What makes conscious epiphenomenalism implausible is the dubiousness of supposing that the elimination of consciousness would make no difference at all to how things go in the physical world—that the elimination of phenomenal pain, for example, might make no difference to our pain behaviour. But *this* consequence—call it conscious quepiphenomenalism—still follows from the zombie possibility, even given the possibility of overdetermination. So the physicalist response to the zombist is unaffected.

The second caveat is based on an appeal to the still-mysterious nature of causation. For consciousness to be epiphenomenal it must fail to cause—but exactly *what it is* to be a relatum in a causal relation is a somewhat contested notion. It may be that on some understandings of the nature of causation the possibility of zombies is consistent with the non-epiphenomenalism of consciousness. Consider, for example, a 'Humean' account of causation that allows causal laws to be non–counterfactual-supporting regularities: on this view, the mere fact of covariation between phenomenal consciousness and physical events in the actual world is enough to constitute mental-physical causal laws and thus give consciousness causal status. This would be so even if zombies are possible: i.e., if that covariation fails to hold in other possible worlds that are physically indiscernible from this one (or from the way physicalists take this one to be).

A more sophisticated version of this kind of response is to raise the possibility of panprotopsychism,[7] though a certain amount of fancy footwork is required before it can work on behalf of the zombist. Physical descriptions arguably do not capture the *intrinsic* nature of the things they describe—what it is to be a photon, or to have a certain mass, or to be metallic just is to stand in certain *relations* to other things; to interact with other entities or conditions in particular ways. Causal laws relate things extrinsically. But what if causality itself depends somehow upon the intrinsic nature of the things that are causally related? A kind of panprotopsychism supposes that this is so, and further that the intrinsic nature of everything is protophenomenal. In this way, the physicocausal structure of the world is explained and underwritten by protophenomenality.

To this we can add the following claims. First, phenomenal consciouness is somehow agglomerated out of the intrinsic protophenomenality of brains, and hence inherits the causal relevance of the protophenomenal. From this we can say that consciousness is not epiphenomenal. Second, although the intrinsic nature of things that underpins causality is in the actual world protophenomenal, it need not be: there is at least one other species of intrinsic essence—call it haecc ["hex"]—such that all the protophenomenal could be replaced by

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

haecc leaving the physical structure of the world unchanged. This preserves the zombie possibility.

So, it appears that there are ways of construing the nature of causality such that the zombie possibility is consistent with the non-epiphenomenality of consciousness. The problem with these kind of responses, however, is that they require the zombist to surrender the dialectical high-ground. The initial zombie argument gains its force by granting the physicalist all her (reasonable, non–question-begging) assumptions about the nature of the physical, and then showing that even given all this, the facts about phenomenal consciousness are still undetermined—that is, that no physical facts whatever are sufficient to fix the phenomenal. Among the physicalist assumptions, it is likely, will be certain assumptions about causality: that causal laws are counterfactual supporting, and that physical causation does not depend for its efficacy on the non-physical. In order to launch the kind of response described above the zombist must reject one or more of these assumptions, and thus limit the varieties of physicalism to which the zombie argument applies (or, more accurately, the varieties of physicalism for which, when the zombie attack is made, the epiphenomenalism response is unavailable).

Hence, it is not enough for the zombist simply to point out that non-standard understandings of causation are available on which consciousness need not be epiphenomenal. She must constrain the options for the physicalist by proffering positive arguments to show that causation actually *is* one or other of these ways. This is a heavy dialectical burden to carry, transferring the burden of proof emphatically onto the zombist, and zombists have not so far made effective sallies on this front.[8]

3. A First Complication: The Contingency of Physicalism

Physicalism is a contingent doctrine: the physicalist typically concedes not only that physicalism might not have been true (i.e., that there are possible worlds—even possible worlds that closely resemble our own—in which physicalism does not hold) but also that physicalism *might* not even be true of the actual world (though the physicalist bet is that it is true). This is important since it affects the way in which we must specify the zombie hypothesis; and this in turn gives rise to the first more serious problem for the physicalist intent on convicting the zombist of epiphenomenalism. The normal mode is to specify a zombie world as a world physically indiscernible from our own but in which consciousness is absent. But if physicalism is not true of the actual world, then this formulation will not suffice: there may *be* no possible world physically indiscernible from our own but in which consciousness is absent since, for example, interactionist dualism may just possibly turn out to be true (and assuming that removing consciousness would then alter the causal chains and thus the distribution of the physical).

The proper way to think of the zombie world, thus, is as a possible world specified *according to the physicalist's best guess about the physics of the*

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

actual world but in which phenomenal consciousness is absent. And the proper role of zombie arguments is as part of a destructive dilemma for the physicalist:

a) If physics is (relevantly, including causal closure)[9] as specified by the physicalist, then physicalism must be false because of the zombie possibility.
b) If physics is not (relevantly, including causal closure) as specified by the physicalist, then physicalism is (trivially) false.
c) So physicalism is false.

Seen in this light, the zombie possibility sidesteps commitment to epiphenomenalism in the actual world. That is, if the physicalist is right about physics—such that all physical events, including human behaviour, can be given sufficient explanations in purely physical terms—then (according to the zombie argument) consciousness must in the actual world be non-physical and epiphenomenal. But, consistently with the zombie hypothesis, the physicalist might *not* be right about real-world physics and thus, say, dualist interactionism might be the case—and in that case, consciousness need not be epiphenomenal. That is, interactionist zombists can claim that *because* we are not zombies consciousness in our world is *not* epiphenomenal, while still arguing that the mere possibility of zombie worlds that are physically as physicalists (wrongly) believe our world to be shows that physicalism is false.[10]

One drawback for the zombist, however, is that despite all this the following remains the case: the zombie possibility *plus* commitment to a broadly physicalist account of physics entails the epiphenomenalism (or at least quepiphenomenalism) of consciousness in the actual world. And this is a further commitment—to physicalist physics—that many, if not most, contemporary zombists will wish to make.

Secondly, the weaker conclusion still follows from the zombie possibility that conscious epiphenomenalism must be *possible*, even if not actual. That is: zombie worlds are only a coherent possibility on the condition that if physics were as the physicalist describes, then consciousness would be epiphenomenal. There is a possible world, W, physically rather like this one, and containing the distribution of phenomenal consciousness with which we are familiar in this world, with the important additional stipulation that W is causally closed under physics. Now, from the perspective of W, as it were, physiological zombies are possible: that is, George Bush_{w} might have a zombie twin, even if the real Bush could not. Thus, in this sense, the possibility of physiological zombies is consistent with the non-holding of conscious epiphenomenalism in the actual world. But epiphenomenalism still holds in W. So in any world "with respect to which" physiological zombies are conceivable, consciousness in that world is epiphenomenal. (And if the actual world is such a world, then consciousness is epiphenomenal in this world.)

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

The degree to which this point is pressing for the zombist is open to further debate, but one thing at least is clear: it does not let her off the hook. If the zombie possibility involved a commitment to the actual epiphenomenalism of consciousness then that would pretty much be a slam-dunk for the physicalist; and any zombist who wants to combine her dualism (or neutral monism, or some other anti-physicalist position) with an endorsement of physics as the correct science of the physical continues to face this rebuttal. If the zombist dodges the commitment to epiphenomenal consciousness by, for example, denying causal closure, he or she nevertheless must concede that the zombie possibility is a coherent attack on physicalism only if phenomenal epiphenomenalism is *possible*. Though less evidently untenable than the doctrine of actual epiphenomenalism, the thesis of the coherence of the notion of phenomenal epiphenomenalism is a major commitment that the zombist must assume and with respect to which he or she is vulnerable. Many physicalists are prepared to argue not just that phenomenal epiphenomenalism is false, but that it is impossible—that the very notion is unintelligible. To the extent that *these* arguments succeed—a matter that I shall not attempt to adjudicate here—zombist appeals to the contingency of physicalism cannot avail them.

4. A Second Complication: Gap-Filling or Gap-Bridging

"An interactionist dualist can accept the possibility of zombies, by accepting the possibility of physically identical worlds in which physical causal gaps go unfilled, or are filled by something other than mental processes" (Chalmers 2004, 184). This brief comment by David Chalmers, although aimed at the interactionist dualist, suggests an additional problem for the *physicalist* intent of convicting the zombist of epiphenomenalism. The proposal, I take it, is simply this. Suppose that consciousness is neither epiphenomenal nor quepiphenomenal—that is, its presence is causally required for the occurrence of certain physical events, such as human pain behaviour. As we have seen, this is prima facie inconsistent with the zombie possibility: if consciousness causes some physical behaviour, and since removing the cause typically also removes the effect, a world in which consciousness is absent could not be physically indiscernible from the actual world. But this is so only as long as we insist that an event B for which A is the sole, sufficient cause would not have occurred if A had not occurred (and if we leave everything else in the physical causal nexus up to that point the same). Chalmers suggests that we might abandon this assumption. Either we could simply suppose that B occurs anyway in the zombie world, *uncaused*; or we could remove A but *replace* it in the zombie world with another equivalent cause that is neither physical (since the zombie world is physically indiscernible from the target world) nor phenomenal (since the zombie world contains no consciousness).

This proposal differs from that about the nature of causality, above, in that it does not explicitly trespass into the sphere of physicalist assumptions about the physical: in the actual world, physics can (apparently) be allowed to operate

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

just as the physicalist supposes it does in all non–question-begging respects.[11] It is thus legitimate for the anti-physicalist to introduce it *after* all the relevant physicalist assumptions have been granted. And, although strange sounding at first, Chalmers' gap-bridging/gap-filling proposal seems on a certain amount of reflection neither metaphysically nor logically impossible. If gap-handling is metaphysically possible so are zombie worlds, consistently with consciousness being non-epiphenomenal. So the zombist can apparently block the physicalist attempt to foist epiphenomenalism upon her.

However, all is not *quite* so clear cut. Consider gap-filling: here the proposal is that conscious causes be replaced in the zombie world by new causes that are neither physical nor phenomenal. But such a proposal is schematic at best—what could the inserted causal events *be* if they are neither material nor spiritual? Since we have no idea at all about the nature of these events—we have no candidate substances of which they might be modes—it is hard to assess whether the possibility being raised here is real or not.

There is another difficulty as well, which is clearest if we consider gap-bridging (i.e., simply allowing the causal gaps to remain, but having the physical events around the gaps occur as usual). Here the problem is that it is not clear that we can introduce causal gaps while leaving the physical exactly the same. This is because, as suggested above, the individuation conditions for physical entities are relational: what makes a photon a photon, as far as the physical sciences are concerned, is the way it is embedded in a set of law-like causal relationships with other entities (moving at the speed of light in all frames of reference, exerting no rest mass, transferring momentum inversely proportional to its wavelength on colliding with matter, and so on). So as not to beg any questions in favour of physicalism, let us not presuppose that these causal relations need be with other physical entities. Nevertheless, the presence of causal gaps threatens to disrupt these relations in zombie worlds.

To give a simplified example, suppose that a certain kind of pyramidal cell activity causes pain sensations, and that these pain sensations in turn cause certain nerve signals to be sent to muscles. The two physical event-types in this example are, like all physical event-types, individuated by their law-like connections to other event-types—in this case, including at least one non-physical event-type. But in the zombie world, this relation no longer exists—"pyramidal cell activity" is not a cause of pain sensations, for example. This suggests that, just as a 'photon' which did not transfer momentum in a particular way to the particles with which it collides would not in fact be a photon, 'pyramidal cell activity' in the zombie world is not really pyramidal cell activity. And in that case, the 'zombie world' is not really a zombie world, as it is not physically indistinguishable from the target world.

In short, these considerations might be taken to suggest that it is still impossible to remove non-epiphenomenal consciousness from the world and leave the physical exactly the same, since removing consciousness changes the causal networks in which the physical is embedded—by, for example,

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

introducing gaps—and, since physical things are individuated by their places in the causal network, changing the network inevitably changes the physical.

But now the implausibility shoe is on the other foot. Although it is certainly reasonable to think that the physical is characterized extrinsically, it is less clear that physical events, states, or entities (particularly non-fundamental ones, like neural states) are individuated by *all* their lawlike relations. Thus, it may be open to the zombist to deny that causing pain sensations is essential to being pyramidal cell activity.

5. Conclusion

The upshot of all of this for the zombist-physicalist debate with respect to epiphenomenalism is fairly clear, but somewhat complex. Some frequently presented responses to the attempt to connect zombies with epiphenomenalism, such as appeals to causal overdetermination or non-standard accounts of the causal relation, on inspection fail to make much impact on the physicalist worry. The contingency of physicalism is a more substantial worry, but the zombist can reorient the zombie argument as part of a destructive dilemma for the physicalist and in this fashion avoid commitment to the epiphenomenalism of consciousness in the actual world. Yet, he or she remains committed still to the *coherence* of conscious epiphenomenalism in a possible world where the physics is as the physicalist supposes—including, potentially, the actual world—and on this front the zombist is vulnerable. Finally, however, to the degree that the notion of causal gaps (or causal agents that are neither physical nor phenomenal) can be made plausible, it might be possible for the zombist to evade even this commitment and sever the link between the zombie hypothesis and phenomenal epiphenomenalism entirely.

This last move will be attractive to—perhaps even necessary for—the zombist, given the widespread wariness that exists, among physicalists and anti-physicalists alike, about phenomenal epiphenomenalism. Yet it has as yet been little explored, and remains highly speculative. Perhaps this is the front on which the next battle of the "zombie wars" is to be fought.[12]

Notes

1 The *Journal of Consciousness Studies* Volume 2, Issue 4 (1995)—a special issue on zombies—provides a representative sampling of zombie articles. Some further moves in the debate include Balog 1999, Block and Stalnaker 1999, Bringsjord 1999, Brueckner 2001, Chalmers 1999, Dennett 2001, Dietrich and Gillies 2001, Dretske 2003, Frankish 2007, Hawthorne 2001, Kirk 1999 and 2005, Latham 2000, Levin 2002, Levine 1998, Lynch 2004, Nagel 1998, Perry 2001, Polger 2000, Siewert 1999, Skokowski 2002, Sommers 2002, Stalnaker 2002, Stoljar 2001 and 2006, Tanney 2004, Thomas 1998, Webster 2006, Worley 2003, and Yablo 1999.

2 Kirk 2005 contains an extended argument of this sort: although Kirk does not think that the zombie hypothesis entails consciousness epiphenomenalism in the actual world—for reasons that I think I raise worries for in the following—he does think

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

it entails the *conceivability* of what he calls "e-qualia" (2005, 40). Joe Levine holds that (absent panpsychism) "only if phenomenal properties (qualia) are physically realized can they be causally relevant to the production of physical effects" (2001, 25) and uses this to motivate a defence of physicalism against the zombie attack. See also Dennett 1991, 401 ff.; Putnam 1999, 93 ff.; Aydede and Güzeldere 2001, 100; Perry 2001, 72 ff.; and Papineau 2002, 21 ff.

3 Thanks to an anonymous referee for another paper (Bailey 2006) who pressed me to think harder about these issues. That other paper discusses the putative paradoxes of the epiphenomenalism of phenomenal consciousness.

4 But not always.

> Another answer to the zombie argument one sometimes hears is that if it were right, then consciousness would become epiphenomenal. If you could have the same behavior without consciousness, then consciousness would not be doing any work. This answer rests on a misunderstanding. The point of the zombie argument is to show that consciousness, on the one hand, and behavior and causal relations, on the other, are different phenomena by showing that it is logically possible to have one without the other. But this logical possibility does not imply that consciousness does not do any work in the real world. Analogously: Gasoline combustion is not the same thing as car movement, because it is conceivable to have one without the other. But the fact that it is logically possible for cars to move without gasoline, or indeed without any fuel at all, does not show that gasoline and other fuels are epiphenomenal (Searle 2004, 103).

Searle's mistake is instructive. In fact, a "zombie car" would indeed be proof of epiphenomenal internal combustion: it would be a car that is physically indiscernible from, say, my Nissan Sentra, and embedded in exactly the same physical context, but whose pistons move in their cylinders even in the absence of ignited gasoline fumes. But "zombie" Sentras are ruled out by physicalism (even if we assume, to play along with Searle's example, that the physical status of gasoline itself is what is in question); if we remove the fuel, while keeping all the (other) physical facts constant, then the engine will simply stop running!

Searle's error is to suppose that *physically different* vehicles that move without fossil fuels are analogues of the zombie case; of course they are not, any more than a physically different organism that produces intelligent behaviour in the absence of consciousness is a zombie. This kind of misreading of the zombie argument is surprisingly common.

5 E.g., Chalmers 1996, chapter 5. See also Harnad 2001; Kim 2005; and the papers by Bill Robinson, Volker Gadenne, Dieter Birnbacher, and Alexander Staudacher in the *Journal of Consciousness Studies* 13, No. 1–2, 2006, a special issue on "Epiphenomenalism: Dead End or Way Out?"

6 Any lawlike connection between consciousness and physics must then, as zombists typically recognise, be a matter of *non*-physical laws that hold over and above, and independently of, the laws captured by physics.

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

What I am calling here "strong epiphenomenalism" differs from the traditional form in that epiphenomena are usually thought to be caused but not themselves to cause. This asymmetric kind of epiphenomenalism is *ruled out* by the possibility of zombies, but this leaves all the interesting questions as yet untouched. (Thanks to Phil Hanson for pointing this out to me.)

Does epiphenomenalism entail the zombie possibility? Robert Kirk has commented that it seems "clear" to him that it does (Kirk 2005, 8 and 39, citing approvingly Stout 1931, 138 ff.). But this is so only if qualia are *non-physical* epiphenomena (otherwise they are fixed by the fixing of the physical and so must be present in the zombie world). Of course, this begs the question against physicalism with respect to consciousness. Kirk considers in a footnote (39, n. 2) the possibility of "a physicalist version of epiphenomenalism," but comments that "[g]iven causal closure, the special properties in question must have physical effects." I cannot see why this must be so: causal closure commits one to the view that all physical events have physical *causes*, not that all physical events have physical *effects*.

7 See, for example, Seager 1995, Chalmers 1996, Rosenberg 2004, Strawson et al. 2006, Skrbina 2007, and the readings in Clarke 2004.

8 But see Rosenberg 2004 for moves along these lines.

9 The principle of causal closure is the following: if a physical event has a cause at *t* then it has a sufficient physical cause at *t*. That is, the physical world is causally self-sufficient, though there may be additional non-physical causes (operating on, for example, non-physical entities) and thus the principle of causal closure does not beg the question in favour of physicalism.

10 An ancestor of this point is the one that was pressed on me by that anonymous referee mentioned previously. For example, Robert Kirk has written that "It is sometimes assumed that the view that zombies are possible entails epiphenomenalism; but that is not so. One may hold that zombies are possible while denying that the actual world is physically closed under causation: one might be an interactionist" (2003).

11 It may, however, violate some more subtle, yet still non–question-begging, physicalist assumptions. For example, this Chalmersian response may assume that it is coherent to suppose that non-physical events can cause physical ones (i.e., that this can happen in other possible worlds), and it is possible that this conflicts with an aspect of the legitimate physicalist understanding of physics. Similarly, the gap-handling response implies that phenomenal events—in the interactionist world supposed to be actual by the gap-handling zombist—do not have causal powers that are counterfactual supporting such that in the absence of consciousness, *ceteris paribus*, there will be an absence of its normal effects. It therefore may be open to the physicalist to deny that such phenomenal events, properly speaking, cause at all, even in an "interactionist" world.

The issue here, recall, is not whether physicalism itself renders zombies impossible—which would be question-begging, since physicalism is just what is in question—but whether the physicalist understanding *of physics* might have indirect implications for the zombie hypothesis (by showing that the zombie possibility entails conscious epiphenomenalism). It is not question-begging to appeal to (reasonable, principled) aspects of the physicalist story about physics

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

since a) the zombist *must* grant these to the physicalist for the sake of argument, in order to show that consciousness outstrips physics thus-understood; and b) most zombists are, in any case, happy to accept the deliverances of contemporary physics about the physical part of the actual world—they typically hold that physics is incomplete, not that it is deeply false.

12 Thanks to audiences at Wilfrid Laurier University, the Canadian Philosophical Association in Toronto, 2006, and the Association for the Scientific Study for Consciousness 2006 at Oxford, for comments on previous versions of this material. This work was supported by a Standard Research Grant from the SSHRC.

References

Aydede, Murat, and Güven Güzeldere

2001 "Consciousness, Conceivability Arguments, and Perspectivalism." *Communication & Cognition*, 34: 99-122.

Bailey, Andrew

2006 "Zombies, Epiphenomenalism, and Physicalist Theories of Consciousness." *Canadian Journal of Philosophy*, 36: 481-510.

Balog, Katalin

1999 "Conceivability, Possibility, and the Mind-Body Problem." *The Philosophical Review*, 108: 497-528.

Block, Ned and Robert Stalnaker

1999 "Conceptual Analysis, Dualism, and the Explanatory Gap." *The Philosophical Review*, 108: 1-46.

Bringsjord, Selmer

1999 "The Zombie Attack on the Computational Conception of Mind." *Philosophy and Phenomenological Research*, 59: 41-69.

Brueckner, Anthony

2001 "Chalmer's Conceivability Argument for Dualism." *Analysis*, 61: 187-93.

Chalmers, David J.

1996 *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

1999 "Materialism and the Metaphysics of Modality." *Philosophy and Phenomenological Research*, 59: 473-96.

2004 "Imagination, Indexicality, and Intensions." *Philosophy and Phenomenological Research*, 68: 182-90.

Clarke, D. S.

2004 *Panpsychism*. Albany: SUNY Press.

Dennett, Daniel C.

1991 *Consciousness Explained*. Boston, MA: Little, Brown.

2001 "The Zombic Hunch: Extinction of an Intuition?" In *Philosophy at the New Millenium*. Edited by Anthony O'Hear. Cambridge University Press, Royal Institute of Philosophy Supplement: 48, pp. 27-43.

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

Dietrich, Eric, and Anthony Gillies
2001 "Consciousness and the Limits of Our Imagination." *Synthèse*, 126: 361-81.
Dretske, Fred
2003 "How Do You Know You Are Not a Zombie?" In *Privileged Access*. Edited by Brie Gertler. Burlington, VT: Ashgate, pp. 114.
Frankish, Keith
2007 "The Anti-Zombie Argument." *The Philosophical Quarterly*, 57: 650-66.
Harnad, Stevan
2001 "No Easy Way Out." *The Sciences*, 41 (2): 36-42.
Hawthorne, John
2001 "Advice for Physicalists." *Philosophical Studies*, 108: 17-52.
Kim, Jaegwon
2005 *Physicalism, Or Something Near Enough*. Princeton: Princeton University Press.
Kirk, Robert
1999 "Why There Couldn't Be Zombies." *Proceedings of the Aristotelian Society*, supp. vol. 73: 1-16.
2003 "Zombies." Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/entries/zombies/.
2005 *Zombies and Consciousness*. Oxford: Clarendon Press.
Latham, Noa
2000 "Chalmers on the Addition of Consciousness to the Physical World." *Philosophical Studies*, 98: 71-97.
Levin, Janet
2002 "Is Conceptual Analysis Needed for the Reduction of Qualitative States?" *Philosophy and Phenomenological Research*, 64: 571-91.
Levine, Joseph
1998 "Conceivability and the Metaphysics of Mind." *Noûs*, 32: 449-480.
2001 *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.
Lynch, Michael P.
2004 "Zombies and the Case of the Phenomenal Pickpocket." *Synthèse*, 149: 37-58.
Nagel, Thomas
1998 "Conceiving the Impossible and the Mind-Body Problem." *Philosophy*, 73: 337-52.
Papineau, David
2002 *Thinking About Consciousness*. Oxford: Oxford University Press.
Perry, John
2001 *Knowledge, Possibility, and Consciousness*. Cambridge, MA: MIT Press.

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

Polger, Thomas
2000 "Zombies Explained." In *Dennett's Philosophy*. Edited by D. Ross, A. Brooks, and D. Thompson. Cambridge, MA: MIT Press, pp. 259-86.
Putnam, Hilary
1999 *The Threefold Cord: Mind, Body, and World*. New York: Columbia University Press.
Rosenberg, Gregg
2004 *A Place for Consciousness: Probing the Deep Structure of the Natural World*. New York: Oxford University Press.
Seager, William
1995 "Consciousness, Information and Panpsychism." *Journal of Consciousness Studies*, 2: 272-88.
Searle, John
2004 *Mind: A Brief Introduction*. Oxford: Oxford University Press.
Siewert, Charles
1999 *The Significance of Consciousness*. Princeton, NJ: Princeton University Press.
Skokowski, Paul
2002 "I, Zombie." *Consciousness and Cognition*, 11: 1-9.
Skrbina, David
2007 *Panpsychism in the West*. Cambridge, MA: MIT Press.
Sommers, Tamler
2002 "Of Zombies, Color Scientists, and Floating Iron Bars." *Psyche*, 8 (22). [http://psyche.cs.monash.edu.au/v8/psyche-8-22-sommers.html]
Stalnaker, Robert
2002 "What Is it Like to Be a Zombie?" In *Conceivability and Possibility*. Edited by Tamar Szabó Gendler. Oxford: Clarendon Press, pp. 385-400.
Stoljar, Daniel
2001 "The Conceivability Argument and Two Conceptions of the Physical." *Philosophical Perspectives*, 15: 394-413.
2006 "Actors and Zombies." In *Content and Modality*. Edited by J. Thomson and A. Byrne. Oxford: Oxford University Press, pp. 1-17.
Stout, G. F.
1931 *Mind and Matter*. Cambridge: Cambridge University Press.
Strawson, Galen, et al.
2006 *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?* Exeter: Imprint Academic.
Tanney, Julia
2004 "On the Conceptual, Psychological, and Moral Status of Zombies, Swamp-Beings, and Other 'Behaviourally Indistinguishable' Creatures." *Philosophy and Phenomenological Research*, 69: 173-86.

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press

Thomas, Nigel
1998 "Zombie Killer." In *Toward a Science of Consciousness II*. Edited by S. Hameroff, A. Kaszniak, and A. Scott. Cambridge, MA: MIT Press, pp. 171-77.
Webster, William R.
2006 "Human Zombies are Metaphysically Impossible." *Synthèse*, 151: 297-310.
Worley, Sara
2003 "Conceivability, Possibility and Physicalism." *Analysis*, 63: 15-23.
Yablo, Stephen
1999 "Concepts and Consciousness." *Philosophy and Phenomenological Research*, 59: 455-63.

https://doi.org/10.1017/S0012217309090076 Published online by Cambridge University Press