

THE ANTI-ZOMBIE ARGUMENT *

Keith Frankish

Kline had mentioned several reports suggesting that a chemical antidote was administered to the zombi victim in the graveyard at the time of his resurrection.... When I asked [Marcel] if he would be able to prepare it for us, he looked momentarily bewildered. Naturally, he replied, one would never make the poison without making the antidote.

(W. Davis, *The Serpent and the Rainbow*, New York: Warner, 1987, p. 111)

In recent years the 'zombie argument' has come to occupy a central role in the case against physicalist views of consciousness, in large part because of the powerful advocacy it has received from David Chalmers.¹ In this paper I seek to neutralize it by showing that a parallel argument can be run for physicalism, an argument turning on the conceivability of what I shall call *anti-zombies*. I shall argue that the result is a stand-off, and that the zombie argument offers no independent reason to reject physicalism.

I. ZOMBIES AND ANTI-ZOMBIES

I begin with a reminder of the zombie argument. The argument is often set out using the framework of two-dimensional semantics, but I shall present it in a simpler version here; nothing will turn on the complications omitted.

In broad terms, physicalism about consciousness is the view that phenomenal properties are not extra features of the world distinct from those that could be catalogued by a completed basic physics (the *microphysical* features).² There are many ways of spelling out this broad characterization, but in all of them it follows that consciousness *supervenies metaphysically* on the microphysical: any world that has the same microphysical laws as ours and the same distribution of microphysical properties will have the same distribution of phenomenal properties too. (This is not,

* NOTICE: This is the author's version of an article published in *The Philosophical Quarterly*, 57(229): 651-66 (2007). It may differ in minor ways from the published version.

¹ D.J. Chalmers, *The Conscious Mind: in Search of a Fundamental Theory* (Oxford UP, 1996); 'Materialism and the Metaphysics of Modality', *Philosophy and Phenomenological Research*, 59 (1999), pp. 473-96; 'Consciousness and its Place in Nature', in D.J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings* (Oxford UP, 2002), pp. 247-72; 'Does Conceivability Entail Possibility?', in T.S. Gendler and J. Hawthorne (eds), *Conceivability and Possibility* (Oxford UP, 2002), pp. 145-200; 'Phenomenal Concepts and the Knowledge Argument', in P. Ludlow, Y. Nagasawa and D. Stoljar (eds), *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument* (MIT Press, 2004), pp. 269-98; 'The Two-Dimensional Argument Against Materialism', in his *The Character of Consciousness* (Oxford UP, forthcoming; pre-publication version available at <http://consc.net/papers/2dargument.html>).

² I assume that a completed basic physics will deal in micro-level laws and entities, but nothing hangs on this assumption. If it should prove false, we might speak instead of the *basic physical* features.

of course, to say that those microphysical features are *necessary* for consciousness: a physicalist in the present sense can allow that there are worlds where consciousness is non-physically realized.) I shall assume that the converse entailment also holds: that if consciousness supervenes metaphysically on the microphysical, then physicalism is true. This is a widespread assumption on both sides of the debate, and in the present context in particular it is unlikely to be challenged. The only view admitting metaphysical supervenience without physicalism is one which posits metaphysically necessary bridging laws linking microphysical properties with distinct phenomenal ones, and this is not a view which defenders of the zombie argument can consistently endorse. (To hold that the laws of nature are metaphysically necessary is, in effect, to deny that whatever is clearly conceivable is metaphysically possible, which is a key premise of the zombie argument.) Thus, if we use the term 'physical' for properties that are either microphysical or metaphysically supervenient on the microphysical, then physicalism about consciousness can be conveniently expressed as the view that consciousness is a physical phenomenon.

Dualism, in contrast, is the view that consciousness is not a physical phenomenon — phenomenal properties are extra features of the world, over and above the microphysical ones, so consciousness does not supervene metaphysically on the microphysical.

I shall call an object *x* a *physical duplicate* of an object *y* if *x* is a duplicate of *y* in all microphysical respects, and hence in all physical ones. Then *zombies* are beings which are physical duplicates of us, inhabiting a world which is a physical duplicate of ours, but lacking consciousness. That is, zombies share all of our physical properties, but are not conscious.³ The zombie argument for dualism is this:

1. Zombies are conceivable
2. If zombies are conceivable, then zombies are metaphysically possible
3. If zombies are metaphysically possible, then consciousness is non-physical
4. So consciousness is non-physical.⁴

The sort of conceivability involved here is, in Chalmers' terms (see 'Does Conceivability Entail Possibility?'), *ideal* and *primary*. A state of affairs is ideally conceivable if its conceivability cannot be defeated by better reasoning. A state of affairs is primarily conceivable if for all we know *a priori* it might *actually* obtain. Primary conceivability is thus an *a priori* matter, constrained only by logic and the meanings of the terms involved. (In contrast, a state of affairs is secondarily conceivable if it conceivably *might have* obtained, given how the world actually is.)

The zombie argument has a good claim to soundness. Premise (1) is plausible. As Chalmers puts it (*The Conscious Mind*, p. 96), his zombie twin is 'just something physically identical to me, but which has no conscious experience — all is dark inside',

³ See, e.g., Chalmers, *The Conscious Mind*, pp. 94-5.

⁴ See, e.g., Chalmers, 'Consciousness and its Place in Nature', p. 249.

and this description seems coherent. (It may be objected that the claim that zombies are conceivable begs the question against interactionist forms of dualism: we can imagine subtracting consciousness while leaving the physical world unchanged only if we assume that consciousness does not affect the physical world. I shall address this point later.)

Premise (2) is an instance of what I shall refer to as the *CP thesis*, the thesis that if a situation is ideally conceivable, then it is metaphysically possible. (Henceforth I shall write simply of possibility; the variety in question will always be the metaphysical one.) The CP thesis is controversial, but a strong case can be made for it. The most important objection to the thesis arises from *a posteriori* necessities. Water is necessarily H₂O, but this is not knowable *a priori*, and it is primarily conceivable that water is something else. It is arguable, however, that primary conceivability always corresponds to *some* genuine possibility, reflecting the *primary intensions* of the concepts involved. (The primary intension of a concept specifies what the concept refers to in the actual world, or in any world considered as actual. It can roughly be thought of as a description of the characteristic features by which we identify the concept's referent.) Thus the primary conceivability of water's not being H₂O corresponds to the genuine possibility that a substance with the identifying features of water (liquid, colourless, odourless, drinkable, etc.) is not H₂O. We are misled only if we wrongly describe this as the possibility that *water* is not H₂O. Moreover, it is arguable that in the case of consciousness we cannot be misled in this way, since anything that possesses the identifying features of consciousness *is* consciousness — or at any rate something just as problematic as consciousness.⁵

Premise (3) is uncontroversial. If it is possible for a creature to possess all our physical properties without possessing our phenomenal properties, then the latter are not among the former.

Physicalist responses to the zombie argument usually attack one or other of premises (1) and (2), the more radical physicalists denying (1), the more conservative ones accepting (1) but denying (2). Chalmers dubs the former 'type-A' materialists, the latter 'type-B'.⁶

I shall adopt a different approach, showing that we can construct an exactly parallel argument for physicalism. I begin by introducing the notion of an *anti-*

⁵ For detailed discussion and defence of the CP thesis, see Chalmers, 'Does Conceivability Entail Possibility?'

⁶ For type-A positions, see, e.g., D.M. Armstrong, *A Materialist Theory of the Mind* (London: Routledge, 1968); D.C. Dennett, *Consciousness Explained* (Boston: Little, Brown, 1991); R. Kirk, *Zombies and Consciousness* (Oxford UP, 2005); D. Lewis, 'What Experience Teaches', in W. Lycan (ed.), *Mind and Cognition* (Oxford: Blackwell, 1990), pp. 499-519. For type-B positions, see, e.g., P. Carruthers, *Phenomenal Consciousness: a Naturalistic Theory* (Cambridge UP, 2000); C.S. Hill, 'Imaginability, Conceivability and the Mind-Body Problem', *Philosophical Studies*, 87 (1997), pp. 61-85; J. Levine, *Purple Haze: the Puzzle of Consciousness* (Oxford UP, 2001); B. Loar, 'Phenomenal States', in N. Block, O. Flanagan, and G. Güzeldere (eds), *The Nature of Consciousness: Philosophical Debates* (MIT Press, 1997), pp. 597-616; W.G. Lycan, *Consciousness and Experience* (MIT Press, 1996); D. Papineau, *Thinking about Consciousness* (Oxford UP, 2002).

zombie. I shall call an object *x* a *bare physical duplicate* of an object *y* if *x* is a physical duplicate of *y* and has no further properties of a non-physical kind. Then we can define anti-zombies as beings which are bare physical duplicates of us, inhabiting a universe which is a bare physical duplicate of ours, but none the less having exactly the same conscious experiences as we do. That is, in the anti-zombie world consciousness is a physical phenomenon, supervening metaphysically on the world's microphysical features — in virtue of token identities, say. (It is worth stressing that the anti-zombie world is one in which consciousness is *metaphysically*, not merely nomologically, supervenient on the microphysical. If the microphysical features were only causally sufficient for the phenomenal ones, consciousness would not be a physical phenomenon in the present sense.) Of course, if physicalists are right, then *we* are anti-zombies, but to a dualist the notion of an anti-zombie will presumably seem as outlandish as that of a zombie does to a physicalist.

The anti-zombie argument for physicalism is this (the notion of conceivability involved is the same as in the zombie argument, and 'possible' means *metaphysically possible*):

5. Anti-zombies are conceivable
6. If anti-zombies are conceivable, then anti-zombies are possible
7. If anti-zombies are possible, then consciousness is physical
8. So consciousness is physical.

Again the argument has a strong claim to soundness. Premise (5) is plausible. My anti-zombie twin is just something which is physically identical with me, and has no non-physical properties, but which *has* conscious experience — the lights are on inside. The description seems as coherent as that of a zombie. (As with the parallel claim about zombies, there is an assumption here that consciousness has no physical effects; I shall discuss this shortly.) It might be objected that there is an asymmetry between (1) and (5), in that the latter requires us to embed a totality-clause ('no further properties of a non-physical kind') under the conceivability-operator, whereas the former does not. However, it is not clear that this makes (5) less plausible, and both premises are on a par to the extent that they both require us to conceive of the absence of something — phenomenal properties in one case, non-physical properties in the other.

The other premises are also plausible. The case for accepting (6) is the same as that for accepting (2), namely, that it is an application of the CP thesis. Premise (7) requires explanation, but, like (3), should be uncontroversial. In the anti-zombie world consciousness is physical, so the microphysical features of that world (the laws and distribution of properties) are metaphysically sufficient for consciousness, and any world with the same microphysical features will have the same distribution of phenomenal properties. But by definition our world has the same microphysical features as the anti-zombie one. Hence the microphysical features of our world are metaphysically sufficient for the existence of consciousness. So the following is true:

9. If anti-zombies are possible, then the microphysical features of our world are metaphysically sufficient for the existence of consciousness.

Given the assumption that metaphysical supervenience entails physicalism, this is equivalent to (7). It might be objected that all this shows is that we have a physical form of consciousness: it does not rule out the possibility that we have a non-physical form *as well*. Perhaps we possess two otherwise identical sets of phenomenal properties, one physical in character, the other non-physical. This worry can be dismissed, however. Even if the suggestion is coherent (and it is not clear that it is), there is absolutely no reason to suppose that it is true, and considerations of simplicity tell overwhelmingly against it.

This, then, is the anti-zombie argument. It may seem strange to claim that the mere conceivability of physicalism entails its truth, but the general strategy is the same as that employed in the zombie argument. In each case the CP thesis is used to move from a psychological claim to a metaphysical one. In the zombie case we imagine subtracting our consciousness while leaving our physical features intact, and conclude that consciousness is non-physical. In the anti-zombie case we imagine subtracting our (putative) non-physical features while leaving our consciousness intact, and conclude that consciousness is physical. If the strategy is sound in the one case, then it is sound in the other.⁷

II. COMMENTS AND AN OBJECTION

I shall say more in defence of the anti-zombie argument shortly. Before doing so, however, I shall add some points of clarification and comparison, and address a possible objection.

First, the conclusion of the anti-zombie argument is that consciousness is *actually* physical, not that it is *necessarily* physical. If sound, the argument establishes that the microphysical features of our world are metaphysically sufficient for consciousness, and thus that consciousness is physical in our world. It does not show that those features are metaphysically necessary for consciousness, and therefore does not rule out functionalist versions of physicalism, according to which consciousness is in fact microphysically realized but could have been realized in other ways.

Secondly, the effect of the argument is to put pressure on the CP thesis. The zombie and anti-zombie arguments cannot both be sound, so if both zombies and anti-zombies are conceivable, then (assuming (3) and (7) are accepted) conceivability does not entail possibility. The argument is similar to one advanced by Stephen Yablo,

⁷ The idea that the zombists' weapons can be turned against themselves is not new, and both Peter Marton and Scott Sturgeon have employed versions of the strategy: see P. Marton, 'Zombies *versus* Materialists: the Battle for Conceivability', *Southwest Philosophy Review*, 14 (1998), pp. 131-8; S. Sturgeon, *Matters of Mind: Consciousness, Reason and Nature* (London: Routledge, 2000), pp. 115-16. However, their arguments, unlike the present one, take a metamodal form.

which turns on the conceivability of a necessarily existing god.⁸ It seems that we can conceive both of the existence of such a being and of its non-existence. But these conceptions cannot both correspond with genuine possibilities, since a necessarily existing being exists in all worlds if it exists in any. Again the CP thesis comes under pressure.

Thirdly, the anti-zombie argument should be distinguished from a metamodal argument for physicalism, outlined by Chalmers (who does not, of course, endorse it). Let p be the conjunction of all microphysical truths about the world and q an arbitrary phenomenal truth. Then

11. It is conceivable that $p \supset q$ is necessary
12. If it is conceivable that $p \supset q$ is necessary, then it is possible that $p \supset q$ is necessary [CP thesis]
13. If it is possible that $p \supset q$ is necessary, then $p \supset q$ is necessary [S5 principle $MLp \supset Lp$]
14. If $p \supset q$ is necessary, then consciousness is physical
15. So consciousness is physical.⁹

(A parallel argument for dualism can be obtained by substituting 'not necessary' for 'necessary' and 'non-physical' for 'physical' throughout.) Closely related metamodal arguments have also been outlined by Peter Marton and Scott Sturgeon.¹⁰

Crucially, this argument differs from the anti-zombie argument in that it involves conceiving of the truth of a modal claim about the space of possible worlds. This feature may be a serious weakness, for it is arguable that the CP thesis applies only to non-modal claims about the distribution of properties within worlds; indeed, since the space of possible worlds is unique, conceiving of different spaces is tantamount to conceiving of the falsity of the CP thesis, as Chalmers points out.¹¹ The anti-zombie argument, in contrast, does not suffer from this weakness. To conceive of anti-zombies, we simply have to imagine a world where the relation between a being's phenomenal properties and its underlying microphysical ones is such that the former are not further properties over and above the latter — for example, where the relation is one of token identity. If it is possible for such a relation to hold, then a modal claim follows, as spelt out in (9), but the claim need not form part of the original conception. Physicalism entails a modal thesis, but need not be conceived as one.

Fourthly, the anti-zombie argument lacks a certain weakness possessed by its opposite number. The problem is that it is arguable that microphysical concepts pick

⁸ S. Yablo, 'Concepts and Consciousness', *Philosophy and Phenomenological Research*, 59 (1999), pp. 455-63.

⁹ Adapted from Chalmers, 'The Two-Dimensional Argument Against Materialism', online pre-publication version, §8.

¹⁰ Marton, 'Zombies versus Materialists: the Battle for Conceivability'; Sturgeon, *Matters of Mind*, pp. 115-16.

¹¹ See Chalmers, 'Materialism and the Metaphysics of Modality' and 'The Two-Dimensional Argument Against Materialism'.

out their referents by their causal roles, rather than by their intrinsic natures. If so, then in conceiving of a microphysical duplicate of our world, we shall be conceiving merely of a *structural* duplicate — a world with functionally identical microphysical properties, distributed in the same way. And the primary conceivability of such a world can show only that consciousness does not supervene on the microphysical structure of our world. It leaves open the possibility that there is an *a posteriori* necessity linking consciousness with the intrinsic nature of the microphysical properties involved — a view which is a form of Russellian monism rather than dualism.¹² No corresponding loophole exists in the anti-zombie argument. Assuming that the CP thesis is sound, the primary conceivability of an anti-zombie world establishes that it is possible for consciousness to supervene metaphysically on the microphysical structure of our world, which in turn establishes that this structure is metaphysically sufficient for consciousness, whatever the intrinsic nature of the microphysical properties involved.

With these points made, I turn to an objection. This is that the anti-zombie argument begs the question against interactionist dualism.¹³ The thought is that we can conceive of a (temporally extended) bare physical duplicate of the actual world only if we assume that the actual world is physically closed under causation. For if some physical events in the actual world are partially determined by facts about a non-physical consciousness, then it is obvious that in a purely physical world with the same microphysical laws and the same initial distribution of microphysical properties, events would unfold differently.¹⁴ At first sight, it might seem that this objection is innocuous, since a very similar one applies to the zombie argument, which also requires us to conceive of a physically identical world without consciousness. Both arguments assume that interactionist dualism is false. This is not the end of the matter, however. Begging the question against interactionist dualism is, arguably, a more serious sin for an anti-dualist argument than for a pro-dualist one. Moreover, the zombie argument can be modified so that it does not beg this question, giving an argument which interactionist dualists could endorse. A world superficially like ours and with the same distribution of phenomenal properties, but physically closed under causation (any causal gaps being filled by physical substitutes for what are actually non-physical processes), I shall call w_{pc} (for *physically closed*). This world w_{pc} might still contain non-physical properties, provided they are epiphenomenal. Beings that are physical duplicates of the inhabitants of w_{pc} but are not conscious are *zombies* _{w_{pc}} . The interactionist dualist can run a modified zombie argument to the conclusion that consciousness is non-physical in w_{pc} . (*Zombies* _{w_{pc}} are conceivable and therefore possible, so the microphysical features of w_{pc} are not metaphysically sufficient for

¹² See Chalmers, *The Conscious Mind*, pp. 134-6, and 'Consciousness and its Place in Nature', pp. 256-7, 265-7.

¹³ I am grateful to Bob Kirk for pressing me to address this objection, and for suggesting the modified version of the zombie argument discussed below.

¹⁴ See J. Perry, *Knowledge, Possibility, and Consciousness* (MIT Press, 2001), pp. 72-7.

consciousness.) Since physicalists will say that w_{pc} is the actual world, it follows that physicalism about consciousness is false.

Similarly, however, *anti-zombies*_{pc} can be defined as beings that are bare physical duplicates of the inhabitants of w_{pc} but none the less have exactly the same conscious experiences as they do. Then a modified anti-zombie argument can support the conclusion that consciousness is physical in w_{pc} . (Anti-zombies_{pc} are conceivable and therefore possible, so the microphysical features of w_{pc} are metaphysically sufficient for consciousness.) This does not entail that physicalism is true, of course; to reach that conclusion would need the additional premise that the actual world is physically closed under causation. However, it does cancel out whatever reason the modified version of the zombie argument provides for thinking that physicalism is false, and interactionist dualists must either refute it or cease to employ that argument.

In what follows I shall focus on the original versions of the two arguments. Since those versions assume that the actual world is physically closed under causation, there is no difference between zombies and zombies_{pc} or between anti-zombies and anti-zombies_{pc}, and the central conceivability issues are the same in both cases.

III. THE MORAL FOR ZOMBISTS

How should advocates of the zombie argument ('zombists') respond to the anti-zombie argument? Since they endorse the CP thesis, they must accept (6), and as I have shown, (7) is uncontroversial. Their only option, then, is to deny (5), that anti-zombies are conceivable. The same conclusion can be reached by a different route. Zombists maintain that zombies are possible. But the possibility of zombies is incompatible with that of anti-zombies. For if there is an anti-zombie world, then its microphysical features are metaphysically sufficient for consciousness. Hence there is no world with the same microphysical features but without consciousness, that is, no zombie world. Another way of making the point is that the unique possible world which is a bare physical duplicate of the actual world (w_{bp} for *bare physical*; this will be the actual world itself, if the actual world has no non-physical features) ought to be a zombie world, if any is. But w_{bp} is also the only candidate for an anti-zombie world. In effect, then, the proponents of the two arguments disagree as to whether the inhabitants of w_{bp} are conscious. If they are not, then w_{bp} is properly described as a zombie world, and zombies are possible; if they are, then w_{bp} is properly described as an anti-zombie world, and anti-zombies are possible. But of course the two sides cannot both be right. Either the inhabitants of w_{bp} are conscious or they are not. Thus the possibility of zombies is incompatible with that of anti-zombies. And if conceivability entails possibility, then it follows that the conceivability of zombies is incompatible with that of anti-zombies.

In short, physicalists have been giving zombists too easy a ride. It seemed that if the CP thesis were sound, all zombists had to do was to show that there is no incoherence in the notion of a zombie. But in fact they also have to show that there *is*

an incoherence in the notion of an anti-zombie. This involves a significant shift of the burden of proof. As Chalmers himself explains:

In general, a certain burden of proof lies on those who claim that a given description is logically *impossible*.... If no reasonable analysis of the terms in question points towards a contradiction, or even makes the existence of a contradiction plausible, then there is a natural assumption in favour of logical possibility.¹⁵

I shall now argue that zombists cannot shoulder this burden, and thus that the zombie argument fails. I shall begin by making the case for the conceivability of anti-zombies, and then deal with some possible counter-arguments. Most of the issues are familiar from the literature, but the present context will cast them in a different light.¹⁶

IV. CONCEIVING OF ANTI-ZOMBIES

As Chalmers notes in 'Does Conceivability Entail Possibility?', conceivability can be either *negative* or *positive*. A state of affairs is negatively conceivable if the hypothesis that it obtains cannot be ruled out *a priori*, i.e., if there is no contradiction latent in it. A state of affairs is positively conceivable if we can imagine a scenario in which the state of affairs holds, and we can flesh out this scenario to an arbitrary degree of detail. *Prima facie*, anti-zombies are conceivable in both senses. To claim that anti-zombies are negatively conceivable is, in effect, to claim that for all we know *a priori*, the phenomenal properties of experience might be physical ones, whether neurological or functional. This claim involves no obvious contradiction. As Locke observed, we cannot rule out *a priori* the possibility that God might, if he pleased, give systems of matter the power to perceive and think.¹⁷

It might be suggested that on analysis, the concept of a phenomenal property turns out to be that of a non-physical property (an essentially subjective one, say), so that it is *a priori* that phenomenal concepts do not apply to anything in the anti-zombie world. This is not persuasive, however. We may have a concept like this (the traditional philosophical concept of *qualia*, perhaps), but physicalists simply deny that consciousness is to be characterized in terms of it. Instead, they define phenomenal properties in a more neutral way, as those properties we refer to when we talk of what

¹⁵ Chalmers, *The Conscious Mind*, p. 96. By 'logically impossible' he means 'inconceivable': see 'Materialism and the Metaphysics of Modality', p. 477.

¹⁶ Peter Marton has also argued in his 'Zombies versus Materialists' that the CP thesis can be used to shift the burden of proof onto the zombist. He points out that if physicalism is treated as a supervenience thesis, then if it is false, it is necessarily false, and hence impossible. But given the CP thesis, if a state of affairs is impossible, then it is inconceivable. Since zombists reject physicalism and endorse the CP thesis, they must therefore show that physicalism is inconceivable.

¹⁷ Locke, *An Essay Concerning Human Understanding*, 5th edn (London: Dent, 1961), Vol. II, pp. 146-7.

our experiences are *like*, and on which we focus when we introspect our experiences. This definition does not preclude physicality.

Nor is it obviously impossible to form a positive conception of how consciousness could be physical. At its simplest, imagining an anti-zombie involves conjoining the ideas of a bare physical duplicate and consciousness, and on the face of it this is no harder than disjoining them in the case of zombies. It is true that there is some imaginative resistance to the idea that consciousness might be physical. 'How could *this*', people sometimes ask, mentally indicating some experience, 'be just a neurological state?'. Difficulty is irrelevant here, however. Conceivability is all or nothing, and one state of affairs may be harder to imagine than another without being less conceivable. (It is, for example, much harder to imagine Ronald Reagan and Freddie Mercury being the same person than to imagine their being distinct, but the two scenarios are on a par with respect to primary conceivability.) Moreover, reluctance to think of consciousness as physical can be explained away by reference to the features of phenomenal concepts. It is generally agreed that these are not physical concepts (that is, they do not represent their objects *as* physical), and that their cognitive role is very different from that of physical concepts (unlike physical concepts, they are deployed in introspection, tend to trigger sensory images, and can be acquired only by those who have had appropriate sensory experiences). And this may suggest that the properties they pick out cannot be physical. But that, of course, would be a fallacy. It remains open that phenomenal concepts pick out neurological or functional states of some kind.

Of course, this needs supplementing with some account of *how* phenomenal concepts could refer to physical properties, but it is not clear that this cannot be done. We need to imagine a creature subjectively just like us and with exactly the same physical structure as ours, but whose phenomenal concepts refer to physical features of its experiences. Exactly what would be required for this depends on the nature of phenomenal concepts and the means by which their reference is determined — things which are matters of considerable dispute. But I can at least sketch a plausible picture. It is widely agreed that phenomenal concepts pick out their referents directly, rather than by way of other properties contingently associated with them.¹⁸ And I can flesh out this view by combining it either with a causal theory of reference, according to which phenomenal concept *c* refers to phenomenal property *P* if there is the right sort of causal dependency between tokenings of *c* and occurrences of *P*, or with a causal-historical theory, according to which *c* refers to *P* if *c*-tokening has been selected to serve as an indicator of *P*. Such accounts are compatible with physicalism — indeed, they are incompatible with anything other than physicalism if the physical world is

¹⁸ See, e.g., Chalmers, 'The Content and Epistemology of Phenomenal Belief', in Q. Smith and A. Jokic (eds), *Consciousness: New Philosophical Perspectives* (Oxford UP, 2003), pp. 220-72; Loar, 'Phenomenal States'; Papineau, *Thinking about Consciousness*; Perry, *Knowledge, Possibility, and Consciousness*; M. Tye, *Consciousness, Color, and Content* (MIT Press, 2000), and 'A Theory of Phenomenal Concepts', in A. O'Hear (ed.), *Minds and Persons* (Cambridge UP, 2003), pp. 91-105.

causally closed. Of course, at present we do not know whether the postulated causal dependencies or selectional histories actually obtain, and so cannot be sure that the corresponding conceptions are coherent. (Empirical information is relevant here, even though we are concerned with primary conceivability, since the question is whether it is conceivable that our *actual physical properties* should be sufficient for consciousness, and to answer this question we need to know what our actual physical properties are.) But equally we do not know that the conceptions, or variants of them, are *not* coherent, so there is no argument here for denying the conceivability of anti-zombies.

V. ZOMBIST RESPONSES

What can zombies say in response to this? Is there some non-obvious incoherence in the description of an anti-zombie world? I can think of three possible lines of argument.

First, zombies may query the coherence of the account of phenomenal concepts just outlined. It is arguable that physicalists should accept the following two claims: (i) phenomenal concepts and microphysical concepts both pick out their referents directly, rather than via associated properties; (ii) it is not *a priori* knowable that phenomenal concepts and microphysical concepts co-refer. As I mentioned, there is a strong case for thinking that phenomenal concepts refer directly, and physicalists will probably maintain that microphysical concepts refer directly too; denying this merely opens the door to Russellian monism. Claim (ii) is also plausible; it does not seem possible to determine that phenomenal concepts co-refer with microphysical ones simply by reflection on the concepts themselves. (It is true that if the anti-zombie argument is sound, there is a non-obvious *a priori* argument for that conclusion, but in the present context it would be question-begging to rely on it.) But then it follows that in order to conceive of an anti-zombie world we have to imagine not only that two sets of directly referring concepts co-refer, but also that it is not *a priori* knowable that they do so. The objection is that this cannot be done coherently. For the only explanation of how we can fail to know that two concepts co-refer is that at least one of them refers indirectly and we are ignorant of empirical facts linking the properties by which its reference is mediated with the referent of the other concept.¹⁹ Thus if two concepts co-refer, then *either* it is *a priori* knowable that they do so *or* it is the case that at least one of them refers indirectly, by way of contingently associated properties. Following Papineau (*Thinking about Consciousness*, p. 92), I shall refer to this as the *transparency thesis*, since it is the thesis that if two directly referring concepts co-refer, then it is transparent that they do so.

The objection touches on a large recent literature, but this will not be crucial to the overall argument of this paper and I shall confine myself to two brief comments. First,

¹⁹ See Chalmers, *The Conscious Mind*, pp. 141-2, and 'Materialism and the Metaphysics of Modality', pp. 487-8. See also M. Lockwood, *Mind, Brain and the Quantum* (Oxford: Blackwell, 1989), pp. 136-7.

ad hominem, the transparency thesis has a consequence which some dualists may find unattractive. For it entails that phenomenal concepts do not co-refer with *any* directly-referring theoretical concepts, including non-physical ones. This conclusion is unlikely to be welcomed by those dualists who hope to develop a science of consciousness which does more than merely tabulate psychophysical correlations (for example, one which seeks to explain consciousness in terms of more basic, proto-phenomenal properties.²⁰ Secondly, and more importantly, if a causal or causal-*cum*-historical theory of reference holds for phenomenal concepts, then the transparency thesis looks very implausible. For in that case the reference of these concepts will depend on causal and/or historical factors to which we have no *a priori* access, and it is very hard to see how it could be *a priori* knowable that they co-refer with certain other directly referring concepts.²¹ It might be replied that this is not a response that advocates of the anti-zombie argument can make, since the transparency thesis is entailed by the CP thesis, which they accept. (The thought is that if the transparency thesis were false, then there would be numerous exceptions to the CP thesis, since it would be conceivable, but impossible, that identities involving directly referring terms should not hold.) This point is irrelevant, however, since we are here concerned with the argumentative burdens of zombists, not those of their opponents. The zombie must show that anti-zombies are inconceivable, regardless of whether the anti-zombie argument is sound.

I think this reply is in essence a good one, though it would require more work to establish the case. I shall not pursue the matter here, however, since ultimately nothing will turn on it. (I shall explain why in a moment.)

The second option for zombists is to insist that since zombies *are* ideally conceivable, it follows, given the CP thesis, that anti-zombies are not. Chalmers (in 'The Two-Dimensional Argument against Materialism') takes this line in response to the metamodal argument for physicalism outlined earlier. The CP thesis is *a priori*, he argues, and it is conceivable that the microphysical facts *p* might hold without any given phenomenal fact *q*; so it is *a priori* that $p \supset q$ is not necessary. Hence it is not ideally conceivable that $p \supset q$ is necessary. (He offers a similar response to Yablo's god argument, arguing that a necessarily existing god is not ideally conceivable.) The problem with this tactic, however, is that it simply pits one conceivability intuition against another. The advocate of the anti-zombie argument can run an exactly parallel argument for the view that zombies are not ideally conceivable. In the cases discussed by Chalmers the issue is clouded by the presence of metamodal claims, but these are absent in the anti-zombie case, and the argument is straightforward. The CP thesis is *a priori*, anti-zombies are conceivable, so it is *a priori* that anti-zombies are possible and that zombies are not (since the possibility of one is incompatible with that of the other, as I showed above). Hence zombies are not ideally conceivable. It might be replied that it is *harder* to conceive of anti-zombies than of zombies, and thus that the

²⁰ See, e.g., Chalmers, *The Conscious Mind*, p. 126.

²¹ Papineau, *Thinking about Consciousness*, pp. 91-3.

zombie intuition should take priority over the anti-zombie one. But as I have already noted, difficulty is irrelevant here, and without some *independent* argument for the inconceivability of anti-zombies, it is hard to see why the anti-zombie intuition should be dismissed in this way. Chalmers himself proposes that neither the conceivability of physicalism nor its inconceivability should be used as a premise in argument, but I see no reason to accept this ban. If the claim that zombies are conceivable is admissible as a premise, then why should the parallel claim about anti-zombies not be admissible too?

A third option for zombists would be to appeal to an *a priori* argument for the falsity of physicalism. The anti-zombie world is one in which physicalism is true, so if it is *a priori* that physicalism is false, then the description of the anti-zombie world is incoherent. But what anti-physicalist argument are zombists to draw on? They cannot invoke the zombie argument itself, on pain of begging the question. The zombie argument is sound only if anti-zombies are not conceivable, so we cannot appeal to the zombie argument in order to establish the inconceivability of anti-zombies. Zombists must therefore appeal to an independent anti-physicalist argument.

There are problems with this move, however. First, it is not clear that there *is* an independent argument against physicalism. Arguably, all the major arguments trade in one way or another on the conceivability of zombies. The key premise in the knowledge argument, for example, is that the phenomenal facts cannot be deduced from the physical ones — which is to say that zombies cannot be ruled out *a priori*, and are thus negatively conceivable. Secondly, and more importantly, the move would render the zombie argument itself redundant. If the only way of defending the zombie argument is with an independent argument for the same conclusion, then the zombie argument confers no additional plausibility on that conclusion.

The point can be made more formally. Let i be the claim that anti-zombies are inconceivable, d the claim that dualism is true, p_1 the conjunction of the premises of the zombie argument, and p_2 the conjunction of the premises of the independent argument for dualism. p_1 entails i , as I showed earlier, and I am supposing for the moment that the only reason we have to believe i is that p_2 entails d and d entails i . But then p_1 will be only as plausible as p_2 . Given that p_2 itself entails d , this means that p_1 cannot be more plausible than d independently is, and so does not confer any additional plausibility on d .

The moral is that zombists need to identify grounds for thinking that anti-zombies are inconceivable, grounds which are (a) independent of the claim that zombies are conceivable, and (b) not themselves sufficient to entail the falsity of physicalism. It is now clear why the first response I considered, involving the transparency thesis, cannot save the zombie argument. For even if sound, it fails condition (b). The argument relies on three claims: (i) phenomenal concepts and microphysical concepts refer directly; (ii) it is not *a priori* knowable that phenomenal concepts and microphysical concepts co-refer; (iii) the transparency thesis is true. But these claims jointly entail that phenomenal concepts do not co-refer with microphysical ones —

that is, that physicalism is false. So the argument does not meet condition (b) and is not available to zombists.

I suspect that a similar problem would beset any other attempt to demonstrate that the notion of an anti-zombie is incoherent. Even if a sound argument could be found, the considerations involved would independently entail the falsity of physicalism, making the zombie argument otiose. If this is right, then the zombie argument is either unsound or redundant, and the zombist's argumentative position is untenable.

VI. ENDORSING THE ARGUMENT?

The anti-zombie argument was conceived as a tactical device to neutralize the zombie argument. Its primary function is to show that the CP thesis is a two-edged sword and should be rejected. The natural conclusion to draw is that both the zombie argument and the anti-zombie argument are flawed, for the same reason. This strategy should be congenial to type-B materialists, who accept that zombies are conceivable while denying that they are possible.

But there is another option for the physicalist. One of the following three claims must be rejected: (i) the CP thesis is true; (ii) zombies are conceivable; (iii) anti-zombies are conceivable. And if one judges that the case for the CP thesis is stronger than the case for the conceivability of zombies, then one might decide to reject (ii) instead of (i), and endorse the anti-zombie argument. This line might be taken by a type-A materialist.

Of course, in endorsing the anti-zombie argument, one would incur a corresponding burden of proof. One would need to produce grounds for thinking that zombies are inconceivable, grounds which are (a') independent of the claim that anti-zombies are conceivable, and (b') not themselves sufficient to entail the falsity of dualism. The prospects for doing this may actually be rather better than those for establishing the same for anti-zombies. *Prima facie*, a promising option would be to develop an argument for the view that phenomenal properties are functional properties, perhaps representational ones. This would entail the impossibility of zombies, which are supposed to be functional but not phenomenal duplicates, without trading on the anti-zombie intuition itself. And since functionalism does not entail physicalism, it would satisfy (b') too. A further argument would be needed to establish that the functional properties in question are in fact physically realized — a need which the anti-zombie argument might supply.

These are only *prima facie* considerations, of course, and in any case it is unlikely that anyone would want to employ the anti-zombie argument in the role suggested, since there is a much simpler inductive argument from functionalism to physicalism. Given that all the other functional properties we know of appear to be physically realized, it is reasonable to infer that phenomenal ones will be, too, if they are functional.

VII. CONCLUSION

Dualists should not be zombists. The zombie argument is an elegant and seductive piece of philosophical argumentation. But the idea that we can determine the nature of consciousness by an exercise of the imagination seems too good to be true, and the fact that we can construct an anti-zombie argument suggests that it is not true. When zombies and anti-zombies meet, they annihilate each other, and in so doing reveal that considerations of conceivability have little role to play in debates about the nature of consciousness.²²

²² I am grateful to David Chalmers, Bob Kirk and anonymous referees for helpful comments on earlier drafts. I have also benefited from advice from Peter Carruthers, Tim Chappell, Derek Matravers, David Papineau and Carolyn Price.