

## A Zombie Argument *Against* Dualism

Kevin Corcoran and Johnny Waldrop

(This is culled from a larger paper on neutral monism)

Dualism is often taken to entail the possibility of philosophical zombies. We believe that if dualism were true (if there are zombic worlds), it would entail causal superfluity and that causal superfluity is something all of us should want to avoid. Therefore, we argue that the possibility of zombic worlds actually provides reason to *reject* dualism. Here is the argument:

P1) Some mental events have physical effects

P2) For all physical effects  $x$ , if  $x$  has a mental cause, then if dualism is true,  $x$  has a *sui generis* non-physical cause that is *PI*-causally superfluous. (More on *PI*-causal superfluity below.)

P3) It isn't the case that for all physical effects  $x$ , if  $x$  has a mental cause then  $x$  has a *sui generis* non-physical cause that is *PI*-causally superfluous.

**Conclusion:** *Therefore*, dualism is false.

In defense of P<sub>1</sub>, one might offer quite mundane cases of mental causation, e.g. my willing for my arm to go up and my arm thereby going up. At a minimum, we take it that most dualists will accept that some mental events have physical effects. Now, by “*PI*-causal superfluity,” here’s what we mean. Something is *PI*-causally superfluous just in case it is causally superfluous relative to some *PI*-accessible world. Let  $W$  be the familiar set of possible worlds. Let *PI* be a relation on  $W$ —like any other modal accessibility relation. *PI* is an equivalence relation on  $W$  of *physical fact invariance*. That is, *PI* divides  $W$  into equivalence classes of worlds that are identical with respect to the physical facts.<sup>1</sup> So to say at a world  $w$  that  $p$  is *PI*-possible is just to say that there is some world  $w'$  that is identical to  $w$  with respect to the physical facts at which  $p$  is true. In saying that some physical effect  $x$  has a cause that is *PI*-causally superfluous we do not mean that  $x$  has a cause that is causally superfluous (with respect to  $x$ ) at some *PI*-accessible world. Rather, for a cause  $y$  to be *PI*-causally superfluous with respect to some effect  $x$  is for there to be some *PI*-accessible

---

<sup>1</sup> It might be better to say that *PI* divides  $W$  into worlds that are physically identical *just so long as there are physical matters of fact at those worlds* to avoid any two worlds being mutually *PI*-accessible merely in virtue of there being no physical matters of fact at either world. Another way of saying this would be to say that *PI* is an equivalence relation of physical fact invariance on some subset  $W_P$  of  $W$  such that for every  $w \in W_P$ , there is at least one physical matter of fact at  $w$ . Another important clarification might be to add that the complement of  $W_P$  in  $W$ , call it  $W_Q$  ( $W_Q = W - W_P$ ), the subset of  $W$  at which there are no physical matters of fact, is such that for every  $w \in W_Q$ ,  $w$  *PI*-accesses itself and *only* itself. That is, *PI*-accessibility is equivalent to identity for those worlds in  $W$  at which there are no physical matters of fact.

world  $w$  such that (at  $w$ )  $x$  has a sufficient cause and (at  $w$ ) *were*  $y$  to be a cause of  $x$ ,  $y$  would be causally superfluous with respect to  $x$ .<sup>2</sup>

Dualists should accept  $P_2$ . In effect the superfluity claims that we require from the dualist are not claims to the effect that all physical events have sufficient physical causes. All that is required is that the dualist grant that for every physical effect, if it has a *mental* cause, that effect *could* have had a sufficient *physical* cause without supplementing or changing how things stand with respect to the physical facts. Moreover, many dualists believe that there are sound arguments for dualism, the soundness of which entails the possibility of philosophical zombies. Zombies are physical duplicates of conscious creatures that have no mental lives—there are no conscious experiences for these zombies. At a zombie world—that is, a world with zombies but no conscious creatures—there is certainly no *mental* causation for any physical effect in the customary sense. But then a zombie world is always *PI*-possible with respect to its mentally endowed counterpart since the physical facts are held constant.<sup>3</sup> Moreover, at the zombie world, call it  $w_z$ , whatever dualistic causes there may be for a given physical effect at some world *PI*-possible with respect to  $z$  are certainly such that, at  $z$ , were they to be causes of that physical effect, they would be superfluous. And so dualism, if it is taken to entail the possibility of zombies (or, even, taken with zombic sympathies), entails that if  $x$  is a physical effect and  $x$  has a mental cause then  $x$  has some non-physical cause that is such that there is some physically identical world such that, at that world, were that cause to be a cause of  $x$ , it would be superfluous. In other words, dualism (zombies, etc.) entails that if  $x$  is a physical effect and  $x$  has a mental cause then  $x$  has some *sui generis* non-physical cause that is *PI*-causally superfluous.<sup>4</sup> So much for  $P_2$ . What of  $P_3$ ?

---

<sup>2</sup> Not  $\exists y (Cyx \ \& \ \diamond Syx)$ . But rather that  $\exists y (Cyx \ \& \ \diamond (\exists z Czx \ \& \ (Cyx \ > \ Syx)))$ . (Where ‘ $\diamond$ ’ is understood to connote *PI*-possibility.)

<sup>3</sup> It might be thought that events that are actions are *essentially* actions and so a physical event caused by an agent’s willing it to be so is never identical to a physical event not so-caused. In this case an event such as an agent’s arm being raised willfully (an action) will never exist in a zombie world, where physically indistinguishable events of arm-raising will be non-actions. The argument, though, can be made using mere descriptions of events in a physical causal-historical nexus. Picking out events non-rigidly in this way allows us to cross-identify events that are indistinguishable *qua* causal-historical facts though numerically distinct for non-physical reasons—if there are any such pairs of events. So disputes about the identity conditions of events will not in general threaten the argument since the events in question need not be identified haecceistically; they need only be identified descriptively across mutually *PI*-accessible pairs of worlds. [XXXXX raised this sort of worry in correspondence.]

<sup>4</sup> Of course, the dualist is at liberty to deny that zombie worlds are possible (Moreland, etc.). So, let the claim be that *many* dualists take zombie worlds to be possible (e.g., Chalmers) and so the argument is against the main and most plausible versions of dualism on offer in the literature.

We ought to accept  $P_3$ . The reason is quite straightforward. We ought to avoid systematic causal overdetermination in our metaphysics. We ought also to avoid any sort of quasi-occasionalism, even in a restricted form, if possible. We ought also to be suspicious of any thesis that entails, of garden variety causal phenomena, that those phenomena routinely violate otherwise *a posteriori* robust and stable generalizations about counterfactuals of causal dependence. If  $P_3$  is false then not all of these can be avoided. Denying  $P_3$  saddles one with either causal overdetermination, quasi-occasionalism, or causal stories that are deviant with respect to the counterfactual dependences of events. Let us illustrate.

Let  $x$  be a physical effect that has a mental cause. Consider a *PI*-accessible zombie world or some other *PI*-accessible world at which  $x$  has a sufficient physical cause. At that world  $x$  has a sufficient physical cause. Moreover, that physical causation is plausibly a function of the causal powers of physical objects and the history of physical events at that world.<sup>5</sup> But these are duplicated in the actual world: the causal powers of physical objects are all the same, and the physical history of the world—the succession of physical events—is invariant between the two worlds.

Now suppose the physical causal powers and history of events are enough to secure a sufficient physical cause for any physical event  $x$  in the actual world (just as a sufficient physical cause is secured for  $x$  in the zombie world).<sup>6</sup> Then  $x$  has a sufficient cause in the actual world. But  $x$  also has a *sui generis*, non-physical cause. Then  $x$  is causally overdetermined. But then all physical effects that have mental causes are causally overdetermined. So if the physical causal powers and history of events are enough to secure a sufficient physical cause for any given physical effect, then all physical effects that have mental causes are overdetermined.

Suppose, on the other hand, that the causal powers of physical objects and the physical history of the world *are not* sufficient (contra the zombie world) to secure a sufficient cause of some physical effect  $x$ . Then a sort of occasionalism results for  $x$ . For on this picture some mental causation of the physical is *sui generis*, non-physical, and floats freely above a series of physical

---

<sup>5</sup> Or some other features of the world comprising some physical facts that account for this physical causation.

Whatever among the physical facts is doing the work can be substituted for talk of powers and histories throughout.

<sup>6</sup> The argument does not depend on any parochial view as to what causal relata are. Events, causal powers, objects, specific features of events, combinations thereof, etc. can all be duplicated in a zombie world corresponding to its mentally-endowed world-mate. Further, as discussed above, duplicate cross-identifications of the kind needed can be made non-haecceitistically and so we need not be concerned about the numerical identity of causal relata—whatever those turn out to be.

events that otherwise *would* stand in some causal connection, supplying physical effects with sufficient physical causes, but in fact do not stand in any such causal relation. The causal powers of physical things and the history of physical events is just as it would be if  $x$  were to have a physical cause. But, by hypothesis, it does not have any such physical cause and its real, efficacious cause is a *sui generis* non-physical cause. So, a sort of occasionalism holds true of  $x$  with respect to its mental cause. But then all such mental causation is described by a sort of occasionalism.

Moreover, let  $e$  be a physical effect and  $c$  its mental cause—by assumption,  $c$  is a *sui generis* non-physical cause and  $e$  lacks a sufficient physical cause (in the actual world). It seems that  $e$ , though an effect of  $c$ , is not causally dependent on  $c$ , in David Lewis' sense.<sup>7</sup> Lewis' analysis states that a possible event  $x$  causally depends on some possible event  $y$  iff *were  $y$  to occur,  $x$  would* and *were  $y$  not to occur,  $x$  wouldn't*. We have as good a reason as any to think that  $e$  does not causally depend on  $c$ . For  $e$  to causally depend on  $c$ , the following counterfactual would have to come out true:

- If  $c$  were not to occur,  $e$  would not occur.

But we are led to believe that there are physical duplicate worlds (zombie worlds, etc.) at which  $c$  does not occur (e.g., for the dualist, some *PI*-accessible zombie world) at which  $e$  occurs. Not only do we have no non-question-begging reason to assume that the counterfactual above is in fact true, we have reason to suspect that it is false. So  $e$ , though an effect of  $c$ , does not causally depend on  $c$ . But then no such physical effect, if it has a mental cause and lacks a sufficient physical cause, causally depends on its mental cause. Whatever one thinks about the relationship between causation and Lewis' analysis of causal dependence, *systematic* failure of causal dependence between causes and events is surely to be avoided.

So, if  $P_3$  is denied, then each physical event, if it has a mental cause, is either overdetermined or else is described by a sort of occasionalism conjoined with a failure of causal dependence. Overdetermination, restricted quasi-occasionalism, and systematic failures of causal dependence ought to be avoided. Therefore,  $P_3$  ought not be denied. Thus, we ought to affirm  $P_3$ .<sup>8</sup>

---

<sup>7</sup> Lewis, David. "Causation." *The Journal of Philosophy* 70(17): 556-567, (1973).

<sup>8</sup> We have assumed that any sort of mishmash of overdetermination and occasionalism is not preferable to either one on its own, and so for simplicity we only discussed the two separately; and we imagine any combination to have been covered in the process.

And since the argument is valid, and we ought to accept all three premises, we ought to accept the conclusion: dualism is false.

The possibility of philosophical zombies is often taken to be an entailment of dualism about the mental. However, the possibility of philosophical zombies entail a host of metaphysical difficulties. For example, they entail that physical effects with mental causes are either causally overdetermined or else require a sort of quasi-occasionalism conjoined with a failure of causal dependence. Overdetermination, restricted quasi-occasionalism, and systematic failures of causal dependence are all metaphysical claims that ought to be avoided. Their presence, therefore, provide compelling reasons to reject dualism about the mental.