Return of the Zombies?

by John Perry

In Simone Gozzano and Christopher S. Hill (eds.), *New Perspectives on Type Identity: The Mental and the Physical*. Cambridge: Cambridge University Press, 2012: 251--263.

§1. INTRODUCTION

Qualia are experiences, or at any rate the properties of experiences that makes it *like something* to have the experiences. They are what Herbert Feigl called "raw feels," Thomas Nagel calls the "subjective characters of experience" and David Chalmers calls "phenomenal states".

I believe that experiences are events in brains, and subjective characters are among their properties. This view is not universally held, but seems to be a point of agreement between many property dualists and many physicalists. I think, in addition, that subjective characters are identical with neural properties, physical properties of experiences in the brain, and thus I hold a version of the type-type identity theory, the topic of this book, for at least some mental phenomena.

In *The Conscious Mind* (Chalmers, 1996) Chalmers argues that we can conceive of a possible world that is (i) physically indiscernible from ours and (ii) contains no qualia. There are no experiences in such a world; or, at least, the experiences in such a world have no qualia; it is not like anything to have them; they are pretty short of what it takes to be an experience. If there is such a possible world, qualia are not physical properties of persons or brains.

In *Knowledge, Possibility and Consciousness* (Perry, 2001), I granted the possibility of worlds with no qualia in which beings otherwise like us behave like us (I call these "Zombie worlds"). Even if we grant this, it is not obvious that any of these is, in addition, physically indiscernible from the actual world (and so is what I call an "I-

Zombie world"). I assumed that physical indiscernibility requires that all the causes of physical events be the same as in our world. Given this causal requirement, a possible world without qualia will either lack all of the physical effects of qualia, or they will caused by something else, or they will be uncaused. In all three cases it will not be physically indiscernible, given the causal requirement. It seemed to me, then, that there is a problem for most of us, who think that our experiences do cause things, and that qualia contribute to this, in conceiving of any possible world that is both a Zombie world and physically indiscernible from the actual world, and so no reason for us to think such a world is possible.

If *epiphenomenalism* is true, however, qualia have no physical effects *in the actual world*. If epiphenomenalism is true, there can be I-Zombie worlds; no effects of qualia need to be missing, since there are none in the actual world.[1] I concluded that Chalmers's argument gives only epiphenomenalists a good reason for concluding (or continuing to maintain) that qualia are not physical properties. I argued, however, that it gives no one else such a reason.

In this essay I reconsider the issue. First I amend my argument, giving it a weaker conclusion. I argue however that the weaker conclusion still leaves the Zombie argument unconvincing for most everyone.[2] I do this in §2-§4.

In the remainder of the essay I discuss a point made by Brie Gertler in her valuable review of *Knowledge, Possibility, and Consciousness*. Gertler claims that my criticism of Chalmers' Zombie Argument misconstrues Chalmers's argument, since Chalmers need not accept the causal requirement:

---

[1] Epiphenomenalism is usually a dualist view. I think a physicalist could be an epiphenomenalist, although physics probably makes the idea of an physical event with no effects whatsoever pretty dubious. To accommodate the physicalist epiphenomenalist, we would have to adjust the definitions of indiscernibility.

[2] I saw the need for the amendments in considering Gertler's review, and in discussing her review and my argument with students at a seminar at Florida State University. The reconstruction of my original argument given below, and the amendments to it, as well as some of the ideas for fixing it up, derive from participants in the Murat Aydede and Gene Witmer seminar on consciousness at the University of Florida in 2002. Particular thanks go to Kelly Trogden, Gene Witmer and Kurt Ludwig. This is not to say that any of these folk will find my revised position completely (or, in some cases, even remotely) plausible

> Perry appears to invoke a stronger similarity requirement [than Chalmers intends]: the relevant zombie world must also approximate actual causal laws (p. 77). This is why he sees the zombie possibility as inconsistent with the actual causal efficacy of even nonphysical sensations. If these were causally efficacious in the actual word, he thinks, then the zombie world would physical differ from ours, since the absence of sensations in that world would alter the course of physical events. However, Chalmers' argument does not require that the causal laws remain the same in the zombie world (Gertler, 2002).

In §3--§6 I accept the validity of Gertler's point about Chalmer's intentions, but argue that without the causal requirement, I-Zombie worlds do not suffice to show that qualia are non-physical. So in the end I defend my original position, as amended.

§2. THE ARGUMENT RECONSTRUCTED

Here is a reconstruction of the original argument from (Perry, 2001) , altered in anticipation of the new points that need to be considered. Let P be a physical state, C be a state it is like something to be in, and A be an agent.

1. Suppose in the actual world A's having C is a cause of some physical event X, in the sense of being a necessary part of some causally sufficient condition Z. [Denial of epiphenomenalism]
2. Suppose further that X's occurrence is not *over-determined*, in the sense that even without C, X would have occurred as the effect of some other combination Z' of completely physical causes that in the actual world also cause X . (See below.)
3. Suppose further that some other completely physical combination of causes Z' is not *pre-empted* by Z, in the sense that if Z' would have caused X had C not occurred. (See below)
4. Suppose a I-Zombie world is possible; call it *w*.
5. C is not instantiated in *w* (definition of a Zombie world).
6. X occurs in *w* (since *w* is physically indiscernible from the actual world)

7. In *w,* X is not even partly caused by A's being C (from 3).

8. But then,

   i) in *w*, X occurs without any cause at all, or with a new immaterial cause, or

   ii) w is governed by different laws that the actual world, and things work out so that in w, X is caused by some combination of physical events without the help of C.

9. For *w* to be physically indiscernible from the real world, all of the causes of physical events, and their physical effects, must be the same as in the real world, except insofar as qualia pre-empted some events from having the effects they would have had, if the qualia had not occurred. [The casual criterion of indiscernibility, slightly amended.]

10. If either 8i) or 8ii) is true, *w* is physically *discernible* from the actual world.

11. So we must reject supposition 1, 2 or 3, or conclude that such an I-zombie world is not possible. (Trogdon, 2002)

§3 EPIPHENOMENALISM ISN'T REQUIRED

Epiphenomenalism is a hypothesis about the actual world. In KP&C, I said that one had to be an epiphenomenalist to be able to conceive of a Chalmers-Zombie world, that is, the conclusion of my argument was that 1 has to be rejected. However, I now think that isn't quite right. One could conceive of a Chalmers-Zombie world, even accepting the causal criterion of indiscernibility, or the slight modification of it in 9, and still allow that conscious events were causes of physical events, so long as, in the actual world, all of the events they cause are over-determined, or all of the events they cause would have occurred anyway, because the conscious events pre-empted other potential physical causes, that are still there, and no longer pre-empted, in the I-Zombie world. Hence, in the revised argument, I include the denials of these suppositions, and conclude that at least one of them must be rejected, for one to suppose that there are I-Zombie worlds.

This does not seem to be a very large concession, however, because it seems that the population of philosophers who reject epiphenomenalism, but believe that all cases of causation by conscious events are cases of over-determination or pre-emption cannot be very large, and so the population of philosophers who should be convinced by the Zombie argument is still pretty small, or at any rate not much larger than it would be if it were limited to epiphenomenalists. This is because it is pretty obvious that many cases of conscious events causing physical events are not cases of over-determination or pre-emption.

Consider the practice of giving anesthetic when filling a tooth. The dentist doesn't want me to feel pain as he drills. But he also doesn't want me to react in the ways would if I were to feel pain. He doesn't want me to stiffen my back, reach for the drill, and curse him as best as I can with his drill and hands in my mouth.

Now it is logically possible that although the pain is a necessary part of the actual total cause of my stiffening, cursing, and striking, that result is over-determined, so that even without the pain I would do so. But if that were the case, giving Novocain, although it would eliminate pain, wouldn't be useful for the other ends the dentist has in mind. I think dentists believe that by eliminating the pain, they will eliminate the behavior the pain causes, not only the stiffening and hitting, but also the verbal behavior of complaining about the pain.

It is logically possible that the means we have discovered to eliminate pain also, in fact, eliminate the other over-determining causes of pain behavior. And, if one believed that was so, and something like that was so in the case of all physical events that are caused by conscious events, one could conceive of a Chalmers-Zombie world, without being an epiphenomenalist. But I see no reason to suppose the world is like that. There is certainly no evidence that it is like that.

Again, it is logically possible, I suppose, that every time a quale has an effect X in the actual world there is a completely physical event or state in the actual world that is pre-empted by the quale. If the quale did not occur, the physical part of the actual cause

would combine with this co-opted state or event to produce the effect X. In such a world, phenomenal events are efficacious, but there are always lurking potential alternative causes for the events they in fact cause.

It seems to me that we can look and see that the actual world is not like that. For example, last week I stubbed my toe and the pain caused me to swear, hop around and do a number of other things. I believe that if I had not felt the pain, I wouldn't have done those things. I remember the situation quite clearly, and I don't think there was anything in the vicinity that would have led to the same effect the pain led to. There simply wasn't a pre-emtped cause, ready to take over in the absence of pain.

Sometimes in discussions of counterfactual analyses of ability, situations such as the following are contemplated. It is very important to a mad scientist that at a certain time I curse and hop around. To ensure this, he has planted a rock in my path on which I am virtually certain to stub my toe. But in case I do not stub my toe, or I do and it does not cause pain, he has a back-up plan. He has implanted electrodes in my brain which gave him control of certain of my bodily functions, and which also tell him if the requisite pain is about to happen. If the pain doesn't occur, he will press a button on his console that will make me curse and hop around.

This is logically possible in one case, and even in all cases. It could be that all the effects of qualia would be caused by something else if the qualia did not occur. However, I do not believe this is actually so, and I doubt that anyone does. If someone does believe the actual world is like this, then, even if they are not epiphenomenalists, and even if they are not attracted to the hypothesis of over-determination, they will be able to follow Chalmers' directions and conceive of an I-Zombie world, at least with the causal requirement slightly weakened. So, Chalmers argument does not, I must again admit, presuppose epiphenomenalism. I cannot say for sure that there are not other wildly implausible empirical views about the actual world whose adherents can follow Chalmer's directions.

Granting all of this, I still maintain that Chalmers' Zombie Argument gives no person of common sense--- which I take to exclude epiphenomenalism as well as the two new hypotheses about the actual world--- any reason to suppose that qualia are not physical states.

§4 WEAKER CRITERIA OF PHYSICAL INDISCERNIBILITY

Now let us turn to Gertler's main criticism, that 9, the causal criterion of physical indiscernibility, is neither intended by Chalmers nor required for the Zombie argument. If we give up the causal criterion, my reply to the Zombie argument clearly doesn't work.

A straightforward and weak definition in line with Gertler's analysis is:

(G) A possible world is physically indiscernible from the actual world just in case it is the same with respect to the distribution of physical properties at all space-time points.

If we replace the causal criterion in 9 with (G), neither of the alternatives 8 are ruled out. That is, even one who eschews epiphenomalism, overdetermination, and pre-emption will have to admit that there can be I-discernible Zombie worlds on criterion (G). These will be `qualia-compensating' worlds: worlds in which the events qualia cause in that actual world just happen for no cause, or are caused by immaterial events that are not qualia, or are caused in accordance with the physical laws of the world, which differ from those in the actual world.

I will argue, however, that such Zombie worlds do not serve their dialectical purpose; that is, their possibility does not show the non-physicality of qualia. I'll call worlds in which there is no cause, or a new immaterial cause that is not a quale, somewhat disparagingly, `miracle worlds'. I turn to them first.

In the actual world I have just taken a bite from a cookie, and I am continuing to eat it in part because of how good it tasted, an event involving a phenomenal property. This property won't be instantiated in a miracle Zombie world *w,* so it can't be part of the cause of my continuing to eat the cookie in that world. Still, I *could* just continue to eat the cookie, as an uncaused event, or an

event caused by some new immaterial cause. These are not ruled out by (G). They would be ruled out by a slightly stronger criterion that rules out miracles:

> (G') A possible world is physically indiscernible from the actual world just in case it is the same with respect to the distribution of physical properties at space-time points and in addition includes no uncaused events not occurring in the actual world, and no immaterial causes not occurring in the actual world.

I think the Zombie-arguer needs a criterion at least as strong as (G'), for the argument to show what it is supposed to show. If we stick with (G), Zombie-like arguments will rule out too many things as being physical.

Suppose, for example, we are arguing about whether the property of being a gene is immaterial, as some people perhaps once thought. You argue that we can conceive of a possible world in which there are no genes, but the distribution of physical properties is the same as it is. So we have physical indiscernibility, by standard (G), but no genes. So genes aren't physical.

I reply that this makes no sense. If there were no genes, there would be no inherited characteristics. You reply that in your world, succeeding generations of plants and animals exhibit characteristics that suggest inheritance, just as they do in the actual world. But in your world they are not inherited through the workings of genes, they simply happen. Or they are caused by some other immaterial force not involving qualia, perhaps commands of God without a sensorium. This doesn't seem like a very convincing argument for the immateriality of genes. The most convincing reason for supposing that genes are physical, before they were identified and observed, is that they have physical effects. You conceive of a world without the genes, but preserve all of their physical effects as uncaused, and argue that we don't miss the genes, and so being a gene must be nonphysical. That's not much of an argument.

Indeed, wouldn't the same argument work as well with anything? Suppose you wanted to argue that the property of being a lima bean is immaterial, because we can conceive of a world which is physically indiscernible from ours, in which nothing has the

property of being a lima bean---or more precisely, no bit of space-time has the property of being lima-bean occupied. I say this is not true, because in such a world, we wouldn't see lima beans, grow lima beans, cook lima beans, hide uneaten lima beans under our mashed potatoes, and all of the other things that depend on lima beans being physical and having physical properties. You reply that in the world in question you are allowed miracles, so that people will seem to see lima beans, grow plants that seem to have lima beans on them, buy bags that seem to be full of lima beans, seem to push uneaten lima beans under their mashed potatoes, and so forth. All the effects lima beans would have were they were there are in the world, but no lima beans. Seeming to eat lima beans is nourishing, even.

We are not very moved to suppose that lima beans are immaterial by such an argument. One reason is surely that we know what lima beans are. We already know what has to go on at a point in space-time for it to be filled by a lima bean, so in a world where that isn't happening, we know there are no lima beans. At one time we didn't have that knowledge with genes. With qualia, arguably, we still don't have knowledge of what has to go on at space-time points for them to exist, and so we don't know what must be missing, if they are not there. If this is the difference the argument depends on, then it is, once more, a dualist argument that proceeds from *not* knowing that something is physical, to knowing that it is *not* physical.

We do know, however, what has to go on at various points in space time for qualia to occur. It is a point of agreement that qualia are properties of things that occur in the brain. We are aware of qualia, even though we don't seem them in the way we do lima beans. We know what has to happen in a brain for there to be qualia; there have to be states like *this* (have or imagine having a suitable sensation).

Suppose we are moved to a black and white room for a day or two. When it is time to leave, we are assured that the world has changed in a certain way; physically it is just the way it always has been, and the same physical processes will occur when we are exposed to colored surfaces. But all of the non-physical effects of exposure to colored surfaces have been eliminated. So, if color qualia are non-physical, we won't have any, and if they are physical, we will. In that case, when we exit the room and look at a tomato

we can tell directly whether we have color qualia or not. If we do, then either they are physical, or the experimenters were wrong in thinking they removed all of the non-physical effects.

Qualia are different from genes, in that we are aware of qualia. If one of the qualia I have in a certain situation goes missing, I will notice. If I suddenly turned color blind, or blind, or deaf, I would notice. So I do know what must go on at a certain region in space-time, for qualia of a certain kind to occur. I am looking at a red rose. I attend to the red quale I am having. *That* is exactly what has to occur, in the brain of a person, for a red quale to occur in that person's brain. I can remember having such a quale. *That* is the sort of experience has to happen in the brain of a person for the quale of red to occur. I can't give you a description in the canonical vocabulary of brain science, but I know what has to be going on in the brain.

So qualia are like genes, in that there is (or was) a dispute about whether that which goes on is physical. And qualia are like lima beans, in that we do know what has go on in a region of space time for them to occur. If a Zombie-like argument doesn't work for genes or lima beans, it doesn't work for qualia either.

The friend of Zombies may say, "Well, there is still an important difference between lima beans and qualia. You know what has to go on in your brain, what property the experiences that occur there have to have, in order for their to be qualia, but, given *how* you are aware of what goes on, you can't prove it is a physical." But the Zombie argument is supposed to show that we *can't* be physicalists without denying qualia. I'm only trying to show that is doesn't show that.[3]

[3] All of this suggests a sort of reverse-zombie argument. Can we imagine a world physically indiscernible from ours, including all of the causal laws, in which there are no non-physical effects of physical processes, and no non-physical properties of brain states? Now, are there qualia in such a world? It seems to me that the answers are yes and yes. That is, a world that is entirely physical and which has qualia is conceivable. It follows that qualia are physical properties. Whether this is a good argument or not, I don't know, but it proves as much as the Zombie argument.

To review the point of this section. Suppose X is a phenomenon, which some people think is physical, and others do not. Both grant that occurrences of X have physical effects. Both grant that X occurrences are quite real, and we can say when they occur, and where they would have to occur if they were physical. Both grant that we do not know the canonical physical descriptions of X occurrences. The physicalists think this is so because we just don't yet know. The others think it is because they have no physical descriptions, since they are not physical phenomena.

Now we consider a possible world in which there are no X-occurrences, but their absence is compensated for. All of the physical effects of X-occurrences happen in this world, uncaused. Both sides agree this is a possible world. But is it physically indiscernible from ours? It is physically indiscernible from ours if X-occurrences are not physical, but it is not physically indiscernible from ours if they are physical. The possibility of such a world gives the physicalist no reason to suppose that X-occurrences are not physical.

So I conclude that criterion (G) is not strong enough for the Zombie-arguers purposes. If we don't rule out miracles to fill in for the causes we remove when we remove qualia, our Zombie-worlds aren't convincing. The Zombie argument needs something at least as strong as (G').

§5. THE MATERIAL WORLD AND THE PHYSICAL WORLD

The leaves, however, the second option, that the laws in the Zombie world are different, and the effects of qualia in our world occur in the Zombie world with physical causes, in accordance with the different laws. This is not ruled out by criterion (G').

First I want to observe that it is at least not obvious we should grant the possibility of such worlds. The world of alarms clocks, drills, teeth, dentists and other people is our familiar world of three-dimensional objects enduring through time, each with its full complement of primary qualities. Much of the physicalism of the eighteenth century was the view that this is what the physical world was

like all the way down to some indivisible atoms, whose primary qualities were the basis of the primary qualities of everything composed of them, as well as of the powers of these objects to cause changes in other things --- the secondary qualities and tertiary qualities of the objects. When we are asked to imagine worlds physically indiscernible from our own, it is natural to start with world in which there are alarm clocks, drills, teeth, dentists and other material objects that are the same size and shape and moving at the same rate and entering into all the same spatial relations at different times as the material objects in the real world. If secondary qualities are dispositions to cause qualia of various sorts in certain sorts of physical systems, these material objects will either lack these powers, or have them but be unable to exercise them, depending on how one looks at it. But the tertiary qualities, the powers to affect other sorts of changes in other objects would remain.

When we look at it this way, it seems that we can accept criterion (G'), allow some of the causal facts to change, so that some events caused in the real world have different causes, in accord with different laws, without losing indiscernibility. The physical properties that occur at each point in space-time are still the same.

It is clear that we developed concepts of the material properties of ordinary things, and the ability to recognize when they occurred, measure them, compare them, and the like, long before we knew much about the laws of nature, beyond the rough and ready constraints of folk physics. But of course, it is not material indiscernibility in this sense that is at issue.

What is required is *physical* indiscernibility. Physicalism is based on modern physics, not on eighteenth century physics. To be physically indiscernible from the real world, our Zombie world will not merely have to be physically indiscernible from the actual world at the level of teeth and dentists, cookies and tongues, nor even at the level of molecules and atoms, but all the way down (and all the way up, and all the way out), in whatever sense physics provides for these dimensions.

It is unclear whether, in the context of real physics, as opposed to folk physics, we can make sense two physically indiscernible systems, with the same properties occurring at the same space-time points (or whatever we have instead of space-time points), but with a different nomological basis. It seems arguable that the laws their occurrences instantiate is an essential part of the basic properties of physical science. One philosopher who has advocated this is Sydney Shoemaker (1984), who argues that the true laws of nature are necessary. If that is true of the physics of the real world, then it is by no mean obvious that we can shift to criterion (G') without being committed to a nomological indiscernibility between *w* and the real world in spite of ourselves. And if that is so, then it is at least not obvious that there is any population for whom the Zombie argument should be persuasive, other than those who believe in epiphenemenalism, overdetermination, or pre-emption.[4]

§6. ALTERNATIVE LAWS

But suppose that there is enough give between the actual laws of physics and physical properties that it is conceivable that criterion (G') can be satisfied, in a world governed by different physical laws. And suppose that these physical laws are such that all of the physical effects of qualia in the actual world occur in just the right times and places in the world, even though there are no qualia, and in such a way that we need not suppose that the effects of qualia in the real world are all over-determined or pre-empted. That is, it is qualia-compensating. So there are (G')-indiscernible Zombie worlds. Does this provide any sort of argument that qualia are not physical?

It seems to me that the Zombie argument so conceived is no more plausible than the version that relied on (G) and miracles. The idea is that the alternative laws of nature could be X-*compensating*, in that the absence of some of X's, real events occur in the real world, is compensated for in the otherwise indiscernible world by having all of the effects of X's caused in another way that doesn't involve X's. Why does this show that X's are

[4] This last section was inspired by comments of Olav Gjelsvik at the Oslo University presentation of a version of this paper and further discussions with Bjorn Rambert, Cartsen Hansen, and others after the paper.

not physical? Insofar as it is possible for there to be qualia-compensating worlds, why could there not be lima-bean compensating worlds, or gene-compensating worlds? The possibility of a world governed by X-compensating laws seems quite independent of whether X's are physical or not. So I can't see that this version of the Zombis argument gives the physicalist who is not an epiphenomenalist, and does not believe that all of the effects of qualia are either pre-empted or over-determined in the actual world, any reason whatsoever to give up the hypothesis that qualia are physical.

If the Zombie argument is to be convincing, I think it needs to rely on a causal criterion of indiscernibility. But if it does, it will only be convincing to those who have a very strange view of the actual world.

§7. Conclusion

The basic idea of the Zombie argument is this. Take our world. Remove all the qualia. If what you have left is physically indiscernible from our world, then in removing the qualia you didn't remove anything physical, and so qualia are not physical.

My original reply was basically this. Qualia either have physical effects or they don't. If they do, by taking away all the qualia, you will take away all their physical effects, or you will be changing the status of those events from effects to uncaused events. So you won't end up with a physically indiscernible world. The argument works only if qualia don't have physical effects, that is, if they are epiphenomena. But qualia aren't epiphenomena, so the argument doesn't work.

I've now conceded that the argument would also work if qualia, although not epiphenomena, were over-determining causes, or if they pre-empted other physical causes. These views are perhaps ever so slightly more plausible than epiphenomenalism; I'm not sure. But they are false, so the argument still doesn't work.

If we change the criterion of physical indiscernibility so that a world can be physically indiscernible even though the physical events in it have different

causes, or no causes, we are allowing for qualia-compensating worlds. In these worlds the absence of qualia is compensated for by their actual-world effects occurring as miracles, or as the effects of physical events that occur in the real world, and have these events as effects due to the operation of different laws. But, I argued, the possibility of worlds just like ours, except that some category of objects have been removed, and all their effects have been compensated for, does not show that objects of that category are non-physical.

REFERENCES

Chalmers, David. *The Conscious Mind*. New York: Oxford University Press, 1996.

Feigl, Herbert. *The Mental and the Physical: The Essay and a Postcript.* Minneapolis: University of Minnesota Press, 1967.

Gertler, Brie. Review of *Knowledge, Possibility and Consciousness*. In *Notre Dame Philosophical Reviews*, 5 January 2002.

Nagel, Thomas. "What Is it Like to Be a Bat?", *Philosophical Review*, pp. 435-50

Perry, John. *Knowledge, Possibility and Consciousness*. MIT, 2001.

Shoemaker, Sydney. *Identity*, *Cause*, *and Mind*. Cambridge: Cambridge University Press, 1984

Trogden, Kelly, "A Second Attempt," Email discussed at Aydede-Witmer seminar, University of Florida, 4 April, 2002.