A note on the definition of physicalism

It seems as if it is conceivable that there is a possible world which is like ours in every physical respect, but in which there are no conscious experiences, and in which there is nothing else there needn't be in order for it to be like ours in every physical respect. The inhabitants of that world, who are traditionally called "zombies", would be physically just like us but phenomenally different – whereas there is something that it is like for me to drink coffee, there is nothing that it is like for the zombie who is physically just like me to drink coffee. The possibility of this world is called, after its inhabitants, the possibility of "zombies".

Since zombies are physically just like us, but phenomenally different, their possibility would show that physicalism is false, since it would show that what it is like to be us is not determined by what we are like physically – physicalism, in other words, is not compatible with the possibility of zombies. For this reason, opponents of physicalism argue from the apparent conceivability to the genuine possibility of zombies, whereas those proponents of physicalism who accept the conceivability of zombies must avoid concluding that zombies are genuinely possible, and argue instead for a gap between conceivability and possibility in this case.

But it's not just the genuine possibility of zombies that would be incompatible with physicalism. It seems to be conceivable, for example, that there is a possible world which is like ours in every physical respect, but in which colour experiences are inverted – so that what it is like to experience a certain colour is replaced with what it is like to experience the complement of that colour. What it is like for inhabitants of this world to experience green, for example, would be what it is like for us to experience red, and *vice versa*. The possibility of this world is called the possibility of "inverted spectra".

Since the inhabitants of a world which is physically just like ours, but in which colour experiences are inverted, are physically just like us, but phenomenally different, their possibility would show that physicalism is false, since it would show that what it is like to be us is not entirely determined by what we are like physically – physicalism, in other words, is incompatible with the possibility of inverted spectra. The possibility

of inverted spectra is important because there is a stronger case for its conceivability than there is for the conceivability of zombies, so it is more difficult for proponents of physicalism to simply deny its conceivability.

It also seems to be conceivable that there is a possible world which is physically just like ours, but in which there are additional nonphysical inhabitants. The nonphysical inhabitants of this world are traditionally called "ghosts", and so the possibility of this world is called, after its inhabitants, the possibility of "ghosts". But it is plausible that the inhabitants of this world which are physically just like us are phenomenally just like us too, and so the possibility of ghosts does not show that what it's like to be us is not determined by what we are like physically. The possibility of ghosts is compatible with physicalism, even though their actual existence is not.

Finally, it seems conceivable that there is a possible world which is physically just like ours, but in which there are additional non-physical entities which have prevented the existence of some conscious experiences which exist in our world – suppose, for example, that a benevolent angel has cast a spell leaving everything physically unchanged, but which stops anybody from feeling pain. Whether this possibility, traditionally called the possibility of "blockers", is compatible with physicalism is contentious: Hawthorne (2002) and Stoljar (2010, 138-9) argue that it is not, whereas Leuenberger (2008) argues that it is. In this paper, we will argue that it is not.

Leuenberger's assertion that physicalism is compatible with the possibility of blockers is a premise in his defence of physicalism from the objection based on the apparent conceivability of zombies. Although we can, according to Leuenberger, conceive of a world physically like ours, but in which conscious experiences are absent, we cannot conceive of such a world in which there is nothing there needn't be in order for it to be physically like ours, but only such a world in which additional nonphysical entities prevent the existence of conscious experiences. In other words, Leuenberger argues, we merely conceive of blockers, and so fail to conceive of zombies.

Leuenberger (2008, 158-60) argues that physicalism is compatible with the possibility of blockers in part on the grounds that according to Jackson's (1998, 12) definition, physicalism is compatible with the possibility of blockers. However, we will argue

that the feature of Jackson's definition of physicalism designed to make it compatible with the possibility of ghosts, although it also makes it compatible with the possibility of blockers, inadvertently makes it compatible with the possibility of inverted spectra as well. We will show that a natural revision to Jackson's definition excludes inverted spectra. But that revision excludes the possibility of blockers too.

Jackson (1998, 11) begins his discussion of the definition of physicalism with the following formulation, which is inconsistent with all four possibilities:

(A) Any two possible worlds that are physical duplicates are duplicates *simpliciter*.

According to this definition, physicalism is not compatible with the possibility of zombies, because if there is a possible world in which all the physical facts are the same as in the actual world, but in which there are no conscious experiences, then there are two possible worlds – that world and the actual world – which are physical duplicates, but not duplicates simpliciter – since that world differs from the actual world by lacking conscious experiences.

Likewise, if there is a possible world in which all the physical facts are the same as ours, but in which all colour experiences are inverted, there are two possible worlds – the actual world and the inverted world – which are physical duplicates but which are not duplicates simpliciter, because the inverted world differs from the actual world in that what it is like to have a green experience in the actual world is switched in the inverted world with what it is like to have a red experience. So formulation (A) is not only incompatible with the possibility of zombies, but is also incompatible with the possibility of inverted spectra.

But (A) is also incompatible with the possibility of ghosts. If there is a possible world in which all the physical facts are the same as in the actual world, but in which there are also additional nonphysical inhabitants (and which is similar to the actual world in all other respects), then there are two possible worlds – the ghost world and the actual world – which are physical duplicates, since the ghost world is one in which all the physical facts are the same as ours, but which are not duplicates simpliciter, because the ghost world differs from the actual world in virtue of the existence of additional nonphysical inhabitants. So (A) is inadequate as a formulation of physicalism.

Finally, (A) is incompatible with the possibility of blockers, for the same reason that it is incompatible with the possibility of ghosts. If there is a possible world in which all the physical facts are the same as in the actual world, but in which a further non-physical entity blocks or prevents the existence of some conscious experiences, then there are two possible worlds – the blocker world and the actual world – which are physical duplicates, but which are not duplicates simpliciter, since the blocker world differs from the actual world in virtue of the non-physical entity and the actual world differs from the blocker world in virtue of some of its conscious experiences.

Jackson (1998, 12) argues that in order to accommodate the possibility of ghosts, the definition of physicalism should be reformulated as follows:

(B) Any world which is a *minimal* physical duplicate of our world is a duplicate *simpliciter* of our world.

A minimal physical duplicate of the actual world, according to Jackson's gloss on this definition, is a world which "… (a) is exactly like our world in every physical respect … and (b) contains nothing else in the sense of nothing more by way of kinds and particulars than it must to satisfy (a)" (Jackson, 1998, 13).

According to this definition, physicalism is not compatible with the possibility of zombies. If there is a possible world in which all the physical facts are the same as in the actual world, but in which there are no conscious experiences (nor anything there need not be for it to be like the actual world in physical respects), then this would be a world which (a) is exactly like our world in every physical respect and (b) contains nothing else in the sense of nothing more than it must to satisfy (a) – since it's part of the stipulation that it is a world in which there is nothing which there need not be for it be like the actual world in physical respects.

But the definition is compatible with ghosts, since although a possible world which is physically like ours, but in which there are additional nonphysical inhabitants, is not a duplicate simpliciter of our world – because it differs from our world in virtue of its additional nonphysical inhabitants – it is also not a minimal physical duplicate of our world – the nonphysical inhabitants which make it not a duplicate simpliciter of our world make it not a minimal physical duplicate. By restricting the thesis to a thesis

about our world, Jackson ensures that according to his definition physicalism is not compatible with the actual existence of ghosts, but is compatible with the existence of ghosts in other possible worlds.

For exactly the same reason as (B) is compatible with the possibility of ghosts, (B) is also compatible with the possibility of blockers (Hawthorne, 2002, 104-5). A possible world in which all the physical facts are the same as in the actual world, but in which an additional nonphysical entity blocks or prevents the existence of some conscious experiences is not a duplicate simpliciter of our world – since it differs from the actual world in virtue of the further nonphysical entity and the actual world differs from it in virtue of the existence of some experiences – but it is not a minimal physical duplicate either – in virtue of that same additional nonphysical entity.

But a world in which all the physical facts are the same as ours but in which all colour experiences are inverted is also not a minimal physical duplicate of the actual world because although (a) that possible world is exactly like our world in every physical respect, (b) that possible world does contain something more than it must in order to satisfy (a). In particular, the inverted world contains colour experiences which it is not the case that it must contain in order to be exactly like the physical world in every physical respect. It is not the case that it must contain the colour experiences it does, because it might have contained the colour experiences of our world instead.

Suppose, for example, that I am having a green experience. Then in a world in which all physical facts are exactly the same as in ours, but in which all colour experiences are inverted, I'm instead having a red experience. This red experience isn't something the inverted world must contain in order to be a physical duplicate of our world, since a world may be a physical duplicate of our world and instead contain my actual green experience. After all, the actual world is a physical duplicate of itself, and it does not contain the red experience. It contains my green experience. So the inverted world is a physical duplicate of the actual world, but not a minimal physical duplicate.

Because the inverted world is not a minimal physical duplicate of our world, it cannot be a counterexample to formulation (B), and so the truth of (B) is compatible with the possibility of inverted spectra. But physicalism is incompatible with the possibility of

inverted spectra, and so the truth of (B) is not sufficient for the truth of physicalism. Suppose, for example, that the zombie world is impossible, and that no other possible world is a minimal physical duplicate of our world. Then (B) would be true. Yet at the same time, we may suppose the inverted world is possible and physicalism is false. So the truth of (B) is not sufficient for the truth of physicalism.

One clarification. In the discussion above, we relied on Jackson's gloss of a minimal physical duplicate of our world, but in other discussions a minimal physical duplicate of our world is glossed as a physical duplicate of our world which is minimal with respect to a partial ordering relation between the physical duplicates of our world (see Leuenberger, 2008, 159-160 and Chalmers, 2012, 151). A physical duplicate of our world is minimal, according to the definition of a minimal physical duplicate in terms of a partial ordering relation, if and only if no other physical duplicate of our world is less than or equal to it (Leuenberger, 2008, 159).

A relation is a partial ordering relation if and only if it is reflexive, antisymmetric and transitive. The relation amongst numbers of being less than or equal to, for example, is a partial ordering relation because every number is less than or equal to itself, if any pair of numbers are less than or equal to each other then they are the same number, and for any triple of numbers such that the first is less than or equal to the second and the second is less than or equal to the third, the first is less than or equal to the third. Since there are many partial ordering relations, definitions of physicalism in terms of a partial ordering relation should define which partial ordering relation is relevant.

A world is less than or equal to a world, according to Leuenberger's attempt to define the relevant partial ordering relation, if and only if all fundamental facts which hold at the first world also hold at the second world (Leuenberger, 2008, 160). According to this definition, Leuenberger argues, our world is less than or equal to a world which is physically like ours but in which additional nonphysical entities have prevented the existence of conscious experience, since the facts about conscious experience are not fundamental facts, and so all the fundamental facts of our world hold at that world. So that world, according to Leuenberger, is not a minimal physical duplicate of our world, and so not a counterexample to (B).

However, if facts about conscious experience are not fundamental, then the possibility of inverted spectra is a counterexample to the antisymmetry of the relation defined by Leuenberger. Because the inverted world and our world differ only in the facts about experience, all fundamental facts of our world are facts of the inverted world, and all fundamental facts of the inverted world are facts of our world and so, according to Leuenberger's definition, our world is less than or equal to the inverted world and the inverted world is less than or equal to our world. Leuenberger, in other words, has failed to define a partial ordering relation at all.

In defence of his attempt to define a partial ordering relation, Leuenberger writes that it is "…antisymmetric given that any two distinct worlds differ in fundamental facts" (Leuenberger, 2008, 160). This suggests Leuenberger may respond to this problem by arguing that facts about experience are fundamental facts, in which case neither is the inverted world less than the actual world, nor vice versa. But if facts about experience are fundamental facts, our world isn't less than or equal to a world in which additional nonphysical entities prevent the existence of conscious experience, since fundamental facts of our world – the facts about conscious experience – don't obtain there.

In that case, either the world in which additional nonphysical entities have prevented the existence of conscious experiences is a minimal physical duplicate of our world, or some other physical duplicate of our world is less than or equal to it. If the former, then a minimal duplicate of our world is not a duplicate of our world simpliciter. But if the latter, then, as long as one of those worlds is such that none of the others is less than or equal to it, as Leuenberger (2008, 160) says we may assume, that world is a minimal physical duplicate of our world but not a duplicate simpliciter of our world, since that world is one in which conscious experiences are absent.

So in either case, there is some minimal physical duplicate of our world which is not a duplicate of our world simpliciter and so physicalism, according to definition (B), is incompatible with the possibility of blockers after all. If Leuenberger maintains that facts about experience are not fundamental facts, then he hasn't succeeded in defining the relevant partial ordering relation, because he hasn't succeeded in defining an anti-symmetric relation. But if Leuenberger maintains experiential facts are fundamental

facts, he has succeeded in defining an anti-symmetric relation, but not one according to which the possibility of blockers is compatible with physicalism.

(A world is less than or equal to a world, according to another definition Leuenberger considers, if and only if every positive fact which holds at the first world holds at the second (Leuenberger, 2008, 160; see also Chalmers (1996, 42), Chalmers and Jackson (2001, 210) and Chalmers (2012, 151-2) for discussions of a similar definition). This definition, according to Leuenberger, agrees with the definition we argue for below that physicalism is incompatible with the possibility of inverted spectra and blockers. For this reason, Leuenberger rejects this definition. We prefer to return below to Jackson's gloss, which we regard as more intuitive than the gloss in terms of positive facts, or in terms of partial ordering relations generally.)

The source of the problem posed by inverted worlds can be appreciated by considering Jackson's argument for the sufficiency of (B) for physicalism. Jackson argues (B) is sufficient for physicalism as follows: "If physicalism is false, our world contains some non-physical nature … But that nature cannot be present in any minimal physical duplicate of our world, as that nature is a non-physical addition to the physical nature of our world. But then any such world is not a duplicate *simpliciter* of our world, and, hence, (B) is false" (Jackson, 1998, 14). So according to Jackson, if (B) is true, then physicalism is true as well, and so (B) is sufficient for physicalism.

In the last step of the quotation, Jackson moves from the claim that any minimal physical duplicate of our world is not a duplicate simpliciter of our world to the claim that (B) is false – that it is not the case any minimal physical duplicate of our world is a duplicate simpliciter of our world. But this argument is invalid because if there is *no* minimal physical duplicate of our world then it's vacuously true that any minimal physical duplicate of our world is not a duplicate simpliciter of our world. It does not follow that (B) is false, because in this case it is also vacuously true that any minimal physical duplicate of our world is a duplicate simpliciter of our world.

This is exactly the situation if there is no possibility of zombies, and no other minimal physical duplicate of the actual world is possible, but inverted spectra are possible. In

that situation there is no minimal physical duplicate of the actual world which is not a duplicate of our world simpliciter, simply because there is no minimal physical duplicate of our world at all. There's no minimal physical duplicate of our world at all because any physical duplicate of our world must contain experiences (since zombies are impossible), but need not contain the specific experiences they do (since inverted experiences are possible instead).

So in order to avoid this problem, the definition has to be formulated to say that there is a minimal physical duplicate of our world as follows:

(C) There is a minimal physical duplicate of our world and any world which is a minimal physical duplicate of our world is a duplicate simpliciter of our world.

According to this definition, physicalism is not compatible with the possibility of zombies, because the possibility of a world in which all the physical facts are the same as in our world, but in which there are no conscious experiences, is inconsistent with its second conjunct.

Physicalism is compatible with the possibility of ghosts, according to this definition, because a possible world in which all the physical facts are the same as ours, but in which there are additional nonphysical inhabitants, is not a counterexample to the second conjunct of (C) for the same reason as before – the existence of the same nonphysical inhabitants which make it not a duplicate simpliciter of our world make it not a minimal physical duplicate. And the possibility of ghosts remains compatible with the first conjunct, according to which there is a minimal physical duplicate of our world – the actual world, if it so happens that physicalism is true.

However, definition (C) improves on definition (B) since according to it physicalism is not compatible with the possibility of inverted spectra. Suppose there is a minimal physical duplicate of our world, and there is also a possible world in which all the physical facts are the same as in ours, but in which colour experiences are inverted (and which is similar to our world in all other respects). Then that minimal physical duplicate of our world is not a duplicate simpliciter of our world, because it differs from our world by failing to contain the colour experiences contained in our world. It

fails to contain the colour experiences, because they are more than it requires to be a duplicate of our world.

Suppose, for example, that a minimal physical duplicate of our world duplicates our red experiences. Then if inverted spectra are possible, it contains more than it must in order to be physically exactly like our world, since – as the inverted world shows – it may have been physically just like our world by instead containing green experiences, and so it is not the case that it must contain red experiences. But then it would not be a minimal physical duplicate. So that minimal physical duplicate of our world doesn't contain our red experiences. So that minimal physical duplicate of our world is not a duplicate simpliciter of our world, and the second conjunct of definition (C) is false.

Surprisingly, definition (C) also differs from definition (B) because according to it physicalism is not compatible with the possibility of blockers. Suppose that there is a minimal physical duplicate of our world, and there is also a possible world in which all of the physical facts are exactly the same as in our world, but there is also a further non-physical fact which blocks or prevents the existence of some experiences. Then that minimal physical duplicate of our world is not a duplicate simpliciter of our world, because it differs from our world by failing to contain some experiences. It fails to contain some experiences, because they are more than it must contain to be physically exactly like our world.

Suppose, for example, that a minimal physical duplicate of our world duplicates all of our experiences. Then if blockers are possible, it contains more than it must in order to be physically exactly like our world, because it might have been physically exactly like our world by instead containing a non-physical fact which would prevent or block some of those experiences. But then it would not be a minimal physical duplicate. So that minimal physical duplicate of our world doesn't contain all our experiences. So that minimal physical duplicate of our world isn't a duplicate simpliciter of our world, and the second conjunct of definition (C) is false.

We arrived at (C) by a rather negative path, but we can give a positive argument, in imitation of Jackson's positive argument for (B), for the conclusion that (C) captures physicalism's essential claim. Suppose, to start with, that (C) is false. Then either its

first or second conjunct is false. If the first conjunct is false, then there is no minimal physical duplicate of our world. So if the first conjunct is false, then our world is not a minimal physical duplicate of itself – it contains something more than it must in order to be like our world in every physical respect. But this something more must be non-physical, so if the first conjunct of (C) is false, physicalism is false.

But if the second conjunct of (C) is false, not every minimal physical duplicate of our world is a duplicate simpliciter of our world, and so some minimal physical duplicate of our world is not a duplicate simpliciter of our world. Then our world and a minimal physical duplicate of it differ – at least one contains something the other does not. But since a minimal physical duplicate of our world contains nothing more than it must in order to be like our world in all physical respects, our world must contain something more than it must to be like our world in every physical respect. But this something more must be non-physical, so if either conjunct of (C) is false, physicalism is false.

Conversely, if physicalism is false, then (C) is false. If physicalism is false, then our world contains something non-physical – something more than it must in order to be like our world in all physical respects (this step in the argument assumes, as Jackson's does too, that non-physical entities which must exist, such as God or numbers, are not counterexamples to physicalism). But that something more cannot be present in any minimal physical duplicate of the actual world, so either there is no minimal physical duplicate of our world, or else not every minimal physical duplicate of our world is a duplicate simpliciter of our world, and, hence, (C) is false.[1]

References

Chalmers, David (1996), *The Conscious Mind* (Oxford: Oxford University Press).

Chalmers, David and Frank Jackson (2001), "Conceptual Analysis and Reductive Explanation", *Philosophical Review* 110: 315-61. Reprinted in Chalmers (2010).

Chalmers, David (2010), *The Character of Consciousness* (Oxford: Oxford University Press).

Chalmers, David (2012), *Constructing the World* (Oxford: Oxford University Press).

Hawthorne, John (2002), "Blocking Definitions of Materialism", *Philosophical Studies* 110: 103-113.

Jackson, Frank (1998), *From Metaphysics to Ethics* (Oxford: Oxford University Press).

Leuenberger, Stephan (2008), "Ceteris Paribus Physicalism" in John Hawthorne and Tamar Szabo Gendler (eds.), *Oxford Studies in Metaphysics 4* (Oxford: Oxford University Press).

Stoljar, Daniel (2012), *Physicalism* (London: Routledge).

---

[1] We would like to thank Colleen Gillon, Daniel Stoljar, Peter Tsu and an anonymous referee.