ZOMBIES DEFEATED: A PROJECTIVIST ACCOUNT OF THIRD-PERSON CONSCIOUSNESS ASCRIPTIONS

by

HOWARD J. SIMMONS

July, 2015

I.

A *zombie* in the philosophy of mind is a being that is, by all conceivable tests, indistinguishable from a normal human being, but lacks consciousness. Not all philosophers agree that the concept of a zombie is coherent. (Indeed, I shall add my voice to their doubts below.) Those who do think it coherent use the notion for any of several different purposes. They may argue that since zombies are possible, reductionism of subjective experience to the physical cannot be achieved. (Clearly if it could, any creature like a zombie from a third-person perspective would *ipso facto* have experiences and so not be a zombie.) In addition, they may argue that since zombies are possible, we need to explain why experiences exist at all—why we, for example, are not actually zombies ourselves. There is also the more sceptical thought that since other people might be zombies, I need somehow to justify my belief that they are not.

There has been a great deal of controversy in recent years about whether zombies really are possible. I want here to defend one particular argument against the possibility of zombies, which I think has been unjustly neglected and whose implications will turn out to be rather interesting.

The suggestion that the argument shows zombies to be impossible is not stictly speaking quite accurate. What the argument actually shows is that the zombie hypothesis makes no sense. Those who try to talk about zombies do not actually succeed in talking about anything coherent. So if the argument is sound, the zombie hypothesis is senseless, rather than (necessarily) false.

The argument, due in essence to Lauren Ashwell (Ashwell (2002), 73-81) and Eric Marcus (Marcus (2004)), is as follows. In order to think that *a* is a zombie you have to think both that:

1. *a* is exactly like a human from a third-person perspective;
2. *a* has no experiences.

Thinking (1) is no problem. The difficulty lies in thinking (2), given (1). To see this, note that it is not enough to *not* think that *a* has experiences. That is easy. You can do that just by imagining (1) and forgetting about (2). In other words, you just imagine all the required third-person facts about *a* —essentially what makes *a* 'externally' indistinguishable from a human being—but don't think about whether *a* has experiences or not. But clearly not thinking that *a* has some property is logically distinct from thinking that *a* does not have that property. (To suppose otherwise would be to make an elementary scoping error.) Indeed, when we try to have the latter thought, all we really seem able to manage is the former. But then, assuming that we are not merely subject to a failure of imagination, it is impossible in the deepest sense to think that *a* has no experiences. Hence it is impossible to think that there are, or could be, zombies.

David Chalmers has responded to this argument. He says 'there is no more problem with clearly and distinctly imagining a situation in which there is no consciousness than in imagining a world in which there are no angels or in imagining a world with one particle and nothing else' (Chalmers (2010), 157).[1]

One might take Chalmers' argument in the way he intends it, as a simple refutation of the Ashwell/Marcus argument. On the other hand, one might reasonably think that it compounds the puzzle. How do we imagine a world that lacks angels? As we can see from the logic of the

1 A similarly sceptical response has been provided by Torin Alter (Alter (2007)).

Ashwell/Marcus argument, it is no good just imagining the world with such-and-such features and declining to include angels amongst those features. To suppose otherwise would be to confuse imagining that the world does not contain angels with not imagining that the world contains angels, the same scoping error that we warned against when discussing zombie consciousness.

The trouble with the thought that the world has no angels is that it belongs to a particular category of thoughts whose very meaningfulness can be called into question. These can be referred to as *unrestricted negative existentials* (UNEs). The problem with UNEs is that there is no way to test their truth. If someone were to claim that there are no angels in his back garden, this might be testable. (Any doubts about this would reflect only uncertainty about the concept of an angel and how they might be detected and not the general idea of testing for the existence of something within a limited spacial region.) In contrast how does one test the truth of the claim that there are no angels *in the whole of reality*? What would a test for this even be like?

The claim that a putative zombie *a* (identical to a human in all third-person respects) lacks experiences (and thus really is a zombie) is an unrestricted negative existential, as it is equivalent to the claim that *a*'s experiences do not exist, with no limitation as to the scope of the negated existential. In other words, it means that *a*'s experiences do not exist in the whole of reality. It is thus subject to the same objection raised above against the claim about angels, namely, that it is completely untestable. In general, we can of course test for the existence or non-existence of mental states by observing behaviour, but that is precluded in this case by the very fact that we are being asked to suppose that whether or not *a* has experiences is *not* to be settled by observing *a*'s behaviour.

Some may feel that the foregoing arguments make an unwarrrented assumption of verificationism. That there is some form of verificationism involved here is doubtless correct. But this seems unavoidable when we focus on the task of imagining that *a* has no experiences or imagining that there are no angels. In order to do this our minds need something to focus on. The result of some test or proceess of verification would fulfil this role, but if there is no test that could be done the process of imagining seems impossible.

Is there no way in which unrestricted negative existential thoughts could be legitimised? There is a way, I think, although, as I shall try to show, it does not in fact succeed in saving zombies.

II.

When a particular sort of claim seems philosophically obscure, it can be worthwhile to investigate whether a *projectivist* analysis would be viable. The original and most familiar example is moral utterances. The projectivist about morality denies that moral utterances function essentially by depicting some moral reality. For the projectivist, moral beliefs are *attitudes*.[2] The result is the

2 Some might object that the very idea of a belief presupposes an independent reality to form the subject-matter of that belief. I do not agree with this. I am more inclined to think that the term can be extended to the case of sentences receiving a projectivist treatment (as seems to be the case with terms like 'true' and 'false'). However, if readers prefer to use the word belief in the more restrictive way, then the projectivist strategy can be described differently. Most utterances have what might be called a 'condition of permissible utterance'. So, for example, it is (normally) permissible to utter a paradigmatically factual statement if and only if one has the relevant belief. According to moral expressivism, it is (normally) permissible to utter a moral statement if and only if one has the relevant *attitude*. The projectivist strategy could thus be identified as the task of determining, for given types of utterances, what sorts of conditions of permissibility attach to them, with the assumption that the answer to this question is not going to be that the speaker has certain beliefs.

position known as *moral expressivism*, according to which moral utterances express attitudes. No reference to descriptive or factual content is required.

To be sure, the theory is is not free of difficulties. One of the most persistent concerns the so-called Frege-Geach problem, which, in essence, demands to know how moral sentences can be treated linguistically just *as if* they were factual—in other words, how they can be subject to such operations as negation, conditionalisation, embedding in propositional attitude contexts and so on. Several solutions to this conundrum have been proposed (by, e.g., Gibbard (1990), 83-102; Blackburn (2007)). I am not going to favour any particular technical solution here, but what I do want to do—because I think it will be important for later—is observe that any viable solution must depend in some sense on the fact that moral claims, although they never *mean* anything factual, always have inferential connections with factual claims—this is guaranteed by the familiar point that moral claims supervene on factual ones; in other words, there cannot be a change in a thing's moral properties without a change in some of its factual or 'natural' properties.

What would a projectivist account of UNE claims, including the claim that *a* has no consciousness, look like? Note that a person making this claim is *rejecting* the idea of consciousness as applied to *a*. This point alone could form the basis of a projectivist account. In general, we could say that to believe a claim of the form 'Reality contains no things of type *X*' is to be disposed, for any *Y*, to reject the idea that *Y* is a thing of type *X*. In the case of the zombie claim, we could treat '*a* has no consciousness' as equivalent to 'Reality contains no conscious states of *a*', with the result that believing that *a* has no consciousness is a matter of being disposed, for any object *Y*, to reject the claim that *Y* is a conscious state of *a*. Importantly, we are able, through this device, to distinguish between thinking that *a* has no consciousness and merely not thinking that *a* has consciousness. The distinction lies in the fact that the latter is compatible with indifference to the suggestion, for any *Y*, that *Y* is a conscious state of *a*, whereas the former requires more than just indifference—it needs active rejection.

Does this analysis succeed in saving the zombie idea? Can we use it to make sense of the idea that some creature is a zombie or that we might have been zombies? The answer is no and the next section will explain why.

III.

Consider once more the claim that some creature *a* is a zombie. The projectivist suggestion is that to make this claim is, for any *Y*, to be disposed to reject the idea that *Y* is a conscious state of *a*. But why would anyone want to have this disposition? Remember that zombies are supposed to be exactly like human beings in every respect save the fact that they have no conscious states. So all the evidence favours their *having* conscious states. At this point one might try to push the classic sceptical line, arguing that the evidence is far from conclusive and that *a* might, for all we know, lack any conscious states. However, this would be a wrong move. 'Other minds' scepticism is precluded by the very fact of having adopted a projectivist account. There is no fact of the matter about whether *a* has conscious states. Scepticism argues for a possible mismatch between our beliefs and the facts, but in this stituation there are no facts to start with. (In the same way scepticism about morality cannot get off the ground if one is a moral expressivist.) In any case, even if scepticism were a legitimate stance to take, it would be no help to the person who wants to say that *a* is a zombie, since such a person has to *reject* consciousness on *a*'s part, not merely consider it doubtful.

Strictly speaking, the thesis that zombies are inconceivable does not require us to be able to make sense of the idea that any creature *is* a zombie, only the claim that one or more creatures *might* be zombies (although it would be strange for the modal claim to make sense without the assertoric claim also making sense). But in fact the projectivist account won't allow this either. To see this, contrast the present case with that of moral projectivism. It was noted above that the reason we are able to treat moral utterances as if they were factual even if we are moral projectivists—to negate them, conditionalise them and so on—lies in their inferential connections to factual statements. Another operation that this allows us to do is of course to embed them in modal contexts, to grant as intellegible, for example, not only 'Honesty is good' but also 'Honesty might not have been good'. What does it mean to say that honesty might not have been good? More precisely the question is: what are we meant to imagine when we imagine that honesty is not good? Well, the answer depends on what properties we think something has to have in order to be good. (Of course, on the projectivist account, this is not the same as the *meaning* of 'good'.) Suppose we are consequentialists and think honesty is good because we think that being honest has good consequences. Then if asked to imagine that honesty is not good, we perhaps imagine a world in which people are somewhat different in nature from the way they actually are, with the result that honesty generally does more harm than good. Without some such concrete scenario it seems impossible to get a grip on the idea of honesty's not being good. Now will an account of this kind work for '*a* might be a zombie'? The answer is no and this is because the inferential links between rejection of consciousness for *a* and factual statements about *a* have been broken by the zombie hypothesis itself. These inferential links normally relate third-person statements about a creature—about its behaviour in particular—to ascriptions or non-ascriptions of consciousness to that creature. If they were still intact we could imagine that *a* is not conscious by imagining that *a behaves* like a creature without consciousness. But clearly, given *a*'s indistinguishability in third-person terms from humans, we cannot do this. The conclusion is that it is in fact impossible to make sense of the idea that *a* might a zombie.

IV.

The purpose of this final section is to explore some further implications of what we have discovered.

If the claim that people might have been zombies is meaningless, what follows from this? Well obviously, it follows that this claim cannot be used to undermine mind-to-brain reductionism. (Of course, there may be other ways of undermining this claim that do not require the possibility of zombies: for example, the knowledge argument.) Also, any support that the zombie idea gives to the claim that consciousness needs to be explained or that I need to justify my belief in other minds also evaporates.

But there are also, I think, more positive consequences of the position defended here. In particular, we get an attractive suggestion about how to think about ascriptions of conscious states to beings other than ourselves, one that is invulnerable to sceptical challenges. Instead of thinking of such ascriptions as factual in nature, the idea is that we should think of them as *attitudinal*. To think that a certain creature or type of creature has conscious states is to adopt a certain way of thinking about it, one which involves being prepared to imagine ourselves in that creature's place, trying to feel what that creature is, or might be, feeling.[3] Granted, such attitudinal stances are not arbitrary—they

3 To clarify, I am not suggesting that *all* ascriptions of conscious states to other creatures take this form. (In this respect, the title of this paper is slightly misleading.) A one-off ascription of pain at a particular time to another creature may be treated as factual in nature, true or false according to whether the creature experiences pain at that

are formed on the basis of observing how the creature behaves in different circumstances and deciding whether this behaviour sufficiently matches what we think of as 'amounting to' consciousness. (In this respect, the situation is not like that envisaged in the zombie hypothesis, where the link between ascription or non-ascription of conscious states on the one hand and factual truths about the putative zombie on the other have been broken.) There is a certain slack here. Even when the behaviour of a creature suggests that it has conscious states, it is never irrational not to make the leap and refuse to treat it as conscious. If and when robots whose behaviour is just like that of human beings are created, attributing conscious states to them will be irresistible to many, but it will never be strictly irrational to refuse to do so.[4]

Bibliography

Alter, Torin 2007: "Imagining Subjective Absence: Marcus on Zombies." *Disputatio* 2 (22), pp. 91-101.

Ashwell, Lauren 2002: *Conceivability and Modal Error* (Master's thesis, University of Aukland).

Blackburn, Simon 2007: "Attitudes and Contents" in *Foundations of Ethics: An Anthology*, Russ Shafer-Landau and Terence Cuneo (ed.) (Blackwell), pp. 474-484.

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

Gibbard, Alan 1990: *Wise Choices, Apt Feelings: A theory of Normative Judgement* (Oxford).

.Marcus, Eric 2004: "Why zombies are inconceivable." *Australasian Journal of Philosophy* 82, (3), pp. 477-490.

time. The projectivist account applies to the question of whether a given creature or type of creature has conscious states *in general*.

4 Conversely (but more controversially) it is not irrational, on the perspective here advocated, to treat something like a stone has having conscious states, though it may be very eccentric to do so. You just have to imagine certain feelings and suppose that these are somehow causally related to the physically determinable states of the stone.