Zombie Killer

Nigel J.T. Thomas
86, S. Sierra Madre Boulevard, #5,
Pasadena, CA 91107
nthomas@calstatela.edu

Zombies are hypothetical beings that are behaviorally and functionally (or perhaps even physically) indistinguishable from humans, but which differ from us in not having conscious (or, at least, qualitatively conscious) mental states. Zombiphiles are those who claim that the existence of zombies is a genuine logical possibility, and that this possibility entails that the mind can never be fully understood in functional (i.e. computational), or perhaps even in physical, terms. The 'zombiphile argument', however, only succeeds if the relevant equivalences are understood quite strictly, which can only be made good by the hypothesis of a 'zombie possible-world', identical to the real world but for the fact that each person's 'zombie twin', despite sharing an identical cognitive constitution, environmental situation, and life history, lacks consciousness.

I argue, however, that maintaining the logical possibility of zombies entails consequences that zombiphiles should find unacceptable. *Ex hypothesis*, a zombie makes the same claims about having conscious states as does a normal human. Such claims must either be true, false, or neither.

If they are construed as *false*, the zombie must be understood as either systematically and undetectably lying or as honestly mistaken. Lying may be ruled out on two grounds: (1) ensuring such systematic, undetectable lying would call for differences between the functional architectures of zombies and humans (which is ruled out *ex hypothesis*); (2) beings without a first-hand understanding of consciousness would be unable reliably to identify all situations where lying would be necessary. Standard examples like *Mary the color scientist* and *the inverted spectrum* can be presented without using any 'red flag' terms like "consciousness", "experience" or "qualia" that would alert the zombie to the need to lie; only terms such as "see", "red" and "know", which the zombie would have to be able to use correctly and truthfully in other contexts, are needed.

Since someone's being mistaken normally implies some sort of less-than-optimal cognitive performance, it is not clear in what sense a zombie, which should be construed as cognitively identical to its conscious human 'twin', could be construed as mistaken where the human would not be. Even without assuming such strict cognitive equivalences, if we allowed that zombies could be genuinely mistaken about being conscious, then we could not legitimately exclude the possibility, that our own claims to consciousness might be mistaken. Some philosophers might welcome this conclusion, but it conflicts not only with powerful and widespread intuitions, but with the polemical aims of zombiphiles.

Zombies might be construed as speaking the truth about their conscious experience if the relevant words in their language have subtly different meanings. It has been suggested that instead of attributing *thoughts* to zombies, we should attribute *thoughts*^{z} to them. However, there seems to be little hope that any sense can be made of a notion of *qualia*^{z}. Qualia are supposed to be precisely those entities for which there is no zombie counterpart, yet at least some zombies (like some humans) will claim to have qualia.

If we allow for zombies that are only loosely cognitively equivalent to we humans, and therefore can be envisaged as living in this world alongside us, then zombies might truthfully *admit* their lack of consciousness: they might be (perhaps a subset of) those people (mostly eliminativist philosophers or behaviorist psychologists) who deny its reality. However, this option does not support the zombiphile argument: at best it suggests a program of research aimed at uncovering the functional (or physical) differences underlying the presence or absence of consciousness in its affirmers and its (sincere, non-confused) deniers.

If zombies are neither lying nor telling the truth when they speak about consciousness, what they say must be *empty of meaning*. However, it would not seem to be possible to confine any such emptiness just to their talk about their consciousness. Statements with non-referring terms do not thereby lack truth value ("I have a jaberwock in my pocket," is *false*), and, conversely, we would want some statements by zombies that involve consciousness related terms (e.g. "I *believe* I have qualia") to come out *true*, not meaningless.

It is only possible to make sense of the claim that *all* zombie speech might be meaningless in the light of something like Searle's notion of original/intrinsic intentionality (henceforth *oii*), and the idea is certainly consistent with (but not actually required by) his "connection principle" and "Chinese Room" arguments. Searle might thus be a consistent zombiphile. However, it is highly counter-intuitive to suppose that zombie speech in general (*ex hypothesis* quite as consistently situationally appropriate as that of a human) could be totally empty of meaning. In order to rule out zombies of this type, we may still accept the concept of *oii*, and its close conceptual linkage with consciousness, but should reject the view that human-like behavior is possible in the absence of *oii*. I also reject Searle's view that *oii* is inexplicable in computational/robotic terms. Scientific efforts to understand consciousness would be better served by an initial focus on original/intrinsic intentionality rather than on qualia.