

# Zombies

Intermediate article

Güven Güzeldere, Duke University, Durham, North Carolina, USA

## CONTENTS

*Introduction*  
*History and background*  
*Kinds of zombie*

*Zombie arguments against materialism*  
*Epiphenomenalism and zombies*  
*Conclusion*

*A zombie is a creature that is indistinguishable in behavior as well as in certain physical or physically specifiable respects from a human being, yet which lacks certain mental features that a human being possesses.*

## INTRODUCTION

The folkloric notion of zombies (which came from the West Indies and has been popularized by Hollywood horror movies) attributes life, or at least lifelike behavior, to them, in a resurrected body, while denying them a soul. The philosophical notion postulates the zombie body as identical to a human body characterized at a particular level of abstraction (molecular or functional), and replaces the ‘absence of soul’ with the absence of mind, or particular kinds of mental states. As such, zombies are used as an illustrative element in thought experiments in the philosophy of mind that explore the nature and modality of the relation between bodily properties and mental attributes, as well as the relation between different kinds of mental states (propositional attitudes such as beliefs and thoughts versus conscious experiential states such as visual sensations and pains).

## HISTORY AND BACKGROUND

Because there have been several different ways in which zombies have been described in the philosophical literature as being indistinguishable from humans, and several different features they have been described as lacking, it is necessary to characterize the idea of zombies in general terms. It is possible to come up with at least nine different philosophically relevant notions of a zombie (see Figure 1). However, in most of the notable discussions, zombies are taken to be identical to human beings either in complete physical make-up or in functionally specifiable internal constitution, and in behavior, and they are taken to lack either

conscious qualitative (experiential) mental states or mentality altogether.

Although both the term “zombie” and its core notion had been used in the philosophical literature in a less specific sense earlier (e.g., Martin and Deutscher, 1966; James, 1879, and Campbell, 1970, respectively), it was introduced in its modern form by Robert Kirk (1974) as ‘an organism indistinguishable from a normal human being in all anatomical, behavioral and other observable respects, yet insentient’. Although more general conceptions have since been formulated, Kirk’s discussion, in its aim to refute materialism, not only led the way but also anticipated much of the discussion on zombies that followed in the next three decades. Kirk’s strategy is a familiar one: using a hypothetical argument based on the conceivability of mind, characterized in terms of its cognitive or experiential attributes, and body, characterized in terms of its physical attributes, as having distinct existences. The possibility of body and mind as being thus separable is then used to draw the conclusion that any materialist ontology, however successful in providing an account of the physical world, necessarily falls short of providing a complete account of the mind.

One of the most influential examples of this kind of argument in the history of philosophy is employed in Descartes’s *Meditations*. From the clear and distinct conception of mind, characterized as thinking substance, as existing independently of the body, characterized as extended substance, Descartes concludes that body and mind are in fact ontologically separate and independent. The zombie argument employs the converse of the Cartesian argument. From the self-consistent conception of the body and its full behavioral repertoire as existing in the absence of the mind, it is concluded that a complete theory of bodily attributes and behavior, by itself, is silent about the nature of mind, and the relation between body and mind. It then follows, it is argued, that

		Identity		
		Behavioral	Functional	Physical
Possibility	Natural	(1)	(2)	(3)
	Metaphysical	(4)	(5)	(6)
	Logical	(7)	(8)	(9)

**Figure 1.** The ‘zombie scorecard’: nine distinct notions of zombie, classified according to the respects in which the postulated creature is the same as a human being (the ‘identity’ parameter) and the kind of possible existence the creature is granted (the ‘possibility’ parameter). (Adapted from Polger (2000).)

the accounts provided by materialist theories at best explain the constitution and nature of our zombie twins, not of complete human beings. Lacking the theoretical machinery to talk about the mind, or about how body and mind are related, they fail to distinguish between us and our zombie replicas.

Notice that these two versions of the conceivability argument can be, and often are, proposed independently of one another. Descartes, for example, never employed the ‘zombie argument’ (even though he toys with the idea in his discussions on automata and animals), and, despite the fact that he is generally cited in support of it, it is not obvious that his interactionist substance dualism could indeed permit the metaphysical possibility of a zombie. On the other hand, most anti-materialists today shy away from the Cartesian argument, because of well-known difficulties about the nature of mind–body interaction that have plagued Cartesian theory from the very beginning, and use the zombie argument only in favor of a milder version of dualism, ‘property dualism’.

A few variants of Kirk’s thought experiment have since been proposed as responses to various versions of materialist theories in the philosophy of mind. Most prominently, in the 1980s, the ‘absent qualia’ thought experiment was defended, primarily by Ned Block, contra functionalist materialism. In the 1990s, David Chalmers brought the notion of zombies to bear against materialism of all stripes,

including the identity theory and materialist theories of supervenience. But in terms both of the particulars of the zombie notion employed and of the modal character of the possibility claim, there are important differences between those arguments. In order to better locate these two arguments, I will first sketch a broader framework for discussing zombies.

## KINDS OF ZOMBIE

In general terms, zombies can be classified on the basis of two parameters: physical and mental properties, in virtue of which they are identical to humans in certain respects and different from them in certain others, and the modal strength of their possibility of existence. Güzeldere (1995) explored different kinds of zombies in terms of their postulated constituency, and called them ‘behavioral’, ‘functional’, and ‘physiological’ (or ‘physical’) zombies. Polger (2000) extended the discussion by examining these three kinds of zombies under natural, metaphysical, and logical possibility, producing a  $3 \times 3$  ‘zombie scorecard’ (see Figure 1).

In the first category of identity are creatures that are behaviorally indistinguishable from human beings, but may be made up of completely different, non-carbon-based stuff, with no bodily mechanism, and no functional or computational internal structure, on the basis of which there could be attributed to them a true psychology. They are candidates for ‘behavioral zombies’. It may be that the behavioral zombie goes through the bodily movements that we take to be sophisticated human behavior by a miracle; those movements should not therefore be construed as anything beyond ‘as-if behavior’. Nevertheless, a behavioral zombie is so sophisticated in its mimicry of human behavior that it is, by stipulation, impossible to distinguish it from a normal human being solely on the basis of what (it seems as if) it does. This ‘as-if behavior’ includes, of course, speech acts of the most sophisticated form, which would be sufficient for the behavioral zombie, for example, to pass the Turing Test. However, a behavioral zombie has no internal structure or mechanism that would support a functional description of its psychology, and it would also immediately fail the physical-indistinguishability test once it is internally examined beyond its skin-deep appearance.

In the second category of identity are creatures that are not just indistinguishable from human beings in behavior, but can also be attributed a belief–desire psychology at the right level of

functional characterization. Nonetheless, they may be made not of flesh and bones but of entirely different kinds of matter. This would be the characterization of a 'functional zombie', if it is also postulated that its psychology is incomplete, lacking, in particular, qualitative conscious states. A functional zombie is also a behavioral zombie, but not vice versa.

In the third category of identity are creatures that not only fulfill the criteria of behavioral and functional zombies, but also have the same bodily constituency as human beings, including flesh, blood, bone, nerve cells, and even microtubules – down to the minutest component. We may tentatively regard this kind of creature a candidate for a 'physical zombie'. Of course, a physical zombie is also a functional (and behavioral) zombie. The strongest metaphysical claim based on the possibilities of zombiehood is based on physical zombies.

To recapitulate, at one end of the scale, a behavioral zombie is a creature that is indistinguishable from human beings in its behavior but is unlike a human being in other (physiological and psychological) respects. At the other end, a physical zombie is a replica of a normal human being, identical in all its physical aspects, the psychological attributes of which are, nonetheless, being questioned. A functional zombie lies somewhere in between these two.

A brief examination of which view in the philosophy of mind favors which kind of zombie serves to reveal some of the prior ontological commitments of the various views of the mind. Physicalists, for instance, would need to claim that physical zombies lack nothing at all: that whatever is true of the psychology of humans, including the experiential states and their qualitative phenomenology, will also be true of their physical-zombie counterparts. Functionalists would further assert that functional zombies have a complete mental life, much as we do, because their psychology is functionally equivalent to that of human beings. And metaphysical behaviorists would be committed to the claim that behavioral zombies are just as conscious as any human being, since all mental states are characterizable in purely behavioral and dispositional terms.

Conversely, non-behaviorists, including functionalists and physicalists, may claim that a behavioral zombie would lack crucial aspects of the psychology of a human being; a non-functionalist physicalist may claim that a functional zombie would lack qualia-laden mental states; and property and substance dualists may claim that physical zombies, no matter how perfect molecular replicas they are, can still be 'mindless automata'.

To put the matter differently, for the behaviorist, there are (or can be, in the strongest modal sense) no zombies at all. For the functionalist, the possibility of a behavioral zombie can be admitted, but the possibility of a functional zombie (as well as that of a physical zombie) cannot. And the physicalist has to deny the possibility of a physical zombie.

In addition to the three kinds of zombies distinguished along the identity axis, one can distinguish kinds of zombies along the possibility axis of Figure 1, on the basis of the modal strength of the possibility of their existence. Among the culminating nine elements, I will focus on two particular cells of the zombie scorecard, namely (2) and (9). In recent years, on the basis of the above kind of analysis of zombie kinds vis-à-vis ontological theories, two kinds of zombies in particular, functional and physical, have been used to argue against materialist functionalism under natural possibility and physicalism under logical possibility, respectively.

## ZOMBIE ARGUMENTS AGAINST MATERIALISM

Kirk's zombie argument emerged in the early 1970s at a time when similar ideas were in circulation (e.g. Campbell, 1970; Nagel, 1970; Kripke, 1972), in response to the then-dominant thesis of topic-neutral identity between physical (brain) states and mental states. Campbell's 'imitation man' is an early version of a zombie, and Kripke's contention that God would have additional work to do in order to make the mental properties instantiated after having created and set in place all the physical features of the world evokes, in effect, a complete 'zombie world'.

Materialists, by and large, took this identity relation as contingent, subject to empirical *a posteriori* discovery (Smart, 1959; Armstrong, 1968), although some argued that the identity of brain states and mental states was an analytic truth (Lewis, 1966). Kripke's attack on the identity thesis, based on his theory of rigid designators, was closely related to the zombie-based arguments of the 1970s, which were based on the logical or metaphysical possibility of zombies that were physical replicas of humans but lacked minds altogether.

The debate shifted in the 1980s, as the identity theory by and large gave way to functionalism, and zombie arguments became transformed into arguments about 'absent qualia' and 'ersatz pains'. While the notion of zombies was rarely invoked explicitly during this period, the underlying idea of absent-qualia arguments was the same. Critics of

materialism argued that minds (or mental states) could not be fully characterized in terms of their causal or functional roles, on the basis of thought experiments involving functionally equivalent constructs of brains that, intuitively, did not seem capable of underpinning mental events.

One of the best-known examples of such thought experiments is Ned Block's 'China head' argument for the impossibility of mental states emerging from a brain-like, functionally identical but spatially distributed system (Block, 1978). Block asks us to imagine the functional simulation of a human brain by the Chinese nation, by connecting each of the billion inhabitants of China in appropriate ways through radio links, and having them communicate from a distance like neurons in a brain and thereby animate an artificial body for a certain period of time. According to Block, while this system is 'nomologically possible' and 'could be functionally equivalent to [a human being] for a short time', it is doubtful 'whether it has any mental states at all – especially whether it has "qualitative states", "raw feels", or "immediate phenomenological qualities"'.

The system that Block describes is very much like a functional zombie: it can behave in ways similar to a human being, in virtue of having functionally identical but physically very different internal causal states, and it is Block's contention that it will lack mental states, at least qualitative conscious states. Natural, or nomic, possibility is at issue here. That is, the thought experiment aims to show that there can in fact be systems functionally identical to a human being (or the nervous system of a human being) and that these systems would in fact lack qualitative mental states (cell 2 of the scorecard).

A second type of zombie argument is stronger in its claim: it is based on the logical possibility of physical zombies, and it purports to show that all types of materialist theories are bound to fail (Chalmers, 1996; cell 9 of the scorecard).

Materialists, in response, reject the zombie arguments, either on the basis of differing intuitions on what is logically possible, or by resisting metaphysical conclusions drawn from mere logical possibility claims.

## EPIPHENOMENALISM AND ZOMBIES

Finally, let us examine epiphenomenalism, a view implied by the possibility of zombies. The doctrine of epiphenomenalism has a long history. The philosophers and psychologists of the nineteenth century hotly debated the question of whether

consciousness was part and parcel of the causal network that was responsible for the decisions we make, actions we take, etc., or whether it was just an 'idle spectator', 'riding on' the causal processes, perhaps being caused by them, but without itself exerting any causal force on those processes. Perhaps, some argued, we are all 'automata', since all of our mental life and behavior seems to be determined by our nervous systems, in a purely mechanical framework, with no respectable place in it for consciousness.

This view does not deny that we are conscious. It comes close, however, in positing that consciousness, in itself, makes no difference. Thus it prepares the way for the concept of zombies.

Thomas Huxley was one of the most influential advocates of such a thesis, known as the 'automaton theory of consciousness'. The thesis was first formulated as applying to animals, in agreement with Cartesian intuitions. Huxley (1902) advanced the claim that 'the consciousness of brutes would appear to be related to the mechanism of their body simply as a collateral product of its working, and to be as completely without any power of modifying that working as the steam whistle which accompanies the work of a locomotive engine is without influence upon its machinery'. But the real target was human beings and the nature of human consciousness. This was where Huxley's automata theory differed from Descartes' interactionist dualism. Huxley's account of the 'brutes' was just a lead, to make the same point for humans and maintain that 'in men, as in brutes, there is no proof that any state of consciousness is the cause of change in the motion of the matter of the organism'.

For Huxley, consciousness plays no contributory role in the causal chains in the nervous systems, which totally determine the behavior of an organism; consciousness is merely affected by the neural interactions. In contrast, Descartes' idea of consciousness was of a causally efficacious parameter in the formula of mind-body interaction. Just as Descartes is taken to be the founder of interactionism, Huxley laid a clear foundation for epiphenomenalism with respect to the mind.

## CONCLUSION

Could there be beings who behave like us in every possible way and yet lack consciousness (or possibly all mental life)? Could such beings be not only behaviorally, but also physically identical to us, on a molecular level, and still not have conscious life? On which modal sense of 'could' are we offering our answers – is this a nomic possibility that

can be accommodated within the laws of nature in our world, or is it a metaphysical, or a logical possibility?

The answers one gives to questions of this sort is usually a good indicator of where one stands with respect to a variety of issues regarding consciousness: its ontology, nature, function, evolutionary role, and so on. One's belief in a particular modal possibility of a particular kind of zombie often helps reveal one's implicit metaphysical assumptions, rather than grounding them. As such, the notion of zombies should be considered more of a useful rhetorical tool than the basis of any knock-down argument in philosophy of mind.

## References

- Armstrong D (1968) *A Materialist Theory of the Mind*. London: Routledge.
- Block N (1978) Troubles with Functionalism. In: Savage W (ed.) *Perception and Cognition*, *Minnesota Studies in the Philosophy of Science*, Vol. 9. Minneapolis: University of Minnesota Press.
- Campbell K (1970) *Body and Mind*. New York: Anchor-Books.
- Chalmers D (1996) *The Conscious Mind*. Oxford: Oxford University Press.
- Güzeldere G (1995) Varieties of Zombies. *Journal of Consciousness Studies* 2(4): 326–333.
- Huxley T (1902) *Methods and Results*. New York: Appleton Co.
- James W (1879) Are we automata? *Mind* 4(13): 1–22.
- Kirk R (1974) Zombies v. materialists. *Proceedings of the Aristotelian Society* 48: 135–152.
- Kripke S (1972) *Naming and Necessity*. Cambridge: Harvard University Press.
- Lewis D (1966) An argument for the identity thesis. *Journal of Philosophy* 63(1): 17–25.
- Martin CB and Deutscher M (1966) Remembering. *Philosophical Review* 75: 161–196.
- Nagel T (1970) Armstrong on the Mind. *Philosophical Review* 79(3): 394–403.
- Nagel T (1974) What is it like to be a bat? *Philosophical Review* 83: 435–450.
- Polger T (2000) *Natural Minds*. Ph.D. Dissertation, Duke University.
- Smart JJC (1959) Sensations and brain processes. *Philosophical Review* 68: 141–156.

## Further Reading

- Dennett D (1995) The unimagined preposterousness of zombies. *Journal of Consciousness Studies* 2(4): 322–326.
- Dennett D (1999) *The Zombic Hunch: Extinction of an Intuition?* Royal Institute of Philosophy Millennial Lecture.
- Flanagan O and Polger T (1995) Zombies and the function of consciousness. *Journal of Consciousness Studies* 2(4): 313–321.
- Güzeldere G, Flanagan O and Hardcastle V (1999) The nature and function of consciousness: lessons from blindsight. In: *The New Cognitive Neurosciences*. Gazzaniga M (ed.). Cambridge, MA: MIT Press.
- Kirk R (1999) Why there couldn't be zombies. *Proceedings of the Aristotelian Society*, Supplementary Volume 73: 1–16.
- Levine J (1998) Conceivability and the metaphysics of Mind. *Nous* 32: 449–480.
- Perry J (2001) *Knowledge, Possibility, and Consciousness*. Cambridge, MA: MIT Press.
- Stalnaker R (2002) What is it like to be a zombie? In: Gendler T and Hawthorne J (eds) *Imagination, Conceivability, and Possibility*. Oxford, UK: Oxford University Press.

