

## Zombies and the Function of Consciousness

Owen Flanagan

Departments of Philosophy, Psychology, and Neurobiology,

Email: ojf@acpub.duke.edu.

and

Thomas Polger

Department of Philosophy,

Email:twp2@acpub.duke.edu,

Box 90743

Duke University, Durham, NC, 27707, USA.\*

---

**Abstract:** Todd Moody's Zombie Earth thought experiment is an attempt to show that 'conscious inessentialism' is false or in need of qualification. We defend conscious inessentialism against his criticisms, and argue that zombie thought experiments highlight the need to explain why consciousness evolved and what function(s) it serves. This is the hardest problem in consciousness studies.

---

### 1. Conscious Inessentialism

Conscious inessentialism is "the view that for any intelligent activity *i*, performed in any cognitive domain *d*, even if we do *i* with conscious accompaniments, *i* can in principle be done without these conscious accompaniments."<sup>1</sup> Among other things, conscious inessentialism forces questions about the function, and thus the adaptiveness and evolutionary significance, of consciousness. This is because conscious inessentialism just is the claim that consciousness is not metaphysically or logically necessary. It is

---

\* The order of authors follows alphabetical convention. Authorship is completely equal.

<sup>1</sup> Flanagan, O. *Consciousness Reconsidered*. (Cambridge, MA: MIT Press, 1992), 5.

not necessary in all possible worlds where creatures who are functionally equivalent to us exist. But consciousness did emerge in this actual world—possibly among many species. Why? What adaptive significance does being sentient confer that being terrifically intelligent but insentient does not or could not have conferred under the actual conditions of human evolution?

It is because “conscious inessentialism” is credible that the philosopher’s zombie, the poor sap who is behaviourally indistinguishable from us despite completely lacking phenomenal experience, has figured so prominently in the “consciousness wars.” The zombie is a mere automaton. There is “nothing that it is like” to be a zombie. But zombies can and will fool even the sharpest “mental detector.” Zombies behave just like we do, but are completely “mindless” in the conscious sense.<sup>2</sup>

The zombie problem serves several different purposes. It forces the question of the function of consciousness. It is also a vivid way of illustrating the traditional problem of other minds. How can we be sure that some or all of the people around us are not zombies? And it is often used to show the inadequacy of functionalism, and by implication, of the Turing test.

According to advocates of the orthodox Turing test, if a machine and a person produce the same input-output relations—if they are functionally equivalent—then if we ascribe intelligence to the person, consistency obliges us to also ascribe intelligence to the machine.

---

<sup>2</sup> For a detailed discussion of zombies, zombie worlds, and the philosophical issues surrounding them, see David Chalmers' *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press, 1996.)

According to critics of functionalism, the orthodox Turing test might test for intelligent behaviour, but surely, as the zombie problem shows, it cannot test for consciousness.

Some philosophers have tried to meet this objection by seeking a stronger test than Turing's—call it the “Brainy-Turing-Test.” In addition to behavioural equivalence, the Brainy-Turing-Test demands nervous system equivalence: that is, in addition to input-output equivalence, the brain has to be doing the right things.

The trouble is that even on this more stringent test the conceivability of zombies remains. First, it is possible that there exist creatures who are behaviourally and brain equivalent to us but who are not conscious. For example, when they focus their photoreceptor cells on what we would call a red patch and say, “red,” they are in what is for us the standard brain state of someone “seeing red,” yet there is no experience of seeing red—no experience, period—that supervenes on the allegedly relevant brain state.<sup>3</sup> Second, it is chauvinistic to require nervous system identity conditions for identity of experience. Artificial hearts and kidneys are possible, why not artificial brains (or biologically different ones) that do the same things as the genuine article?

The first point shows that adding in nervous system equivalence is not sufficient to gain experiential equivalence. The second case shows that it is

---

<sup>3</sup> Behavioural and brain equivalence of two organisms where only one of the two is conscious is a possibility so long as nonnaturalism has not been ruled out. If, however, one espouses some version of naturalism according to which mental states are brain states, then one will deny the possibility that there could be two brains in exactly the same state only one of which was conscious.

not necessary. So even the Brainy-Turing-Test is not up to the task of reliably detecting the presence or absence of consciousness.

## 2. Moody Zombies.

In his recent essay “Conversations With Zombies,”<sup>4</sup> Todd Moody puts forward an argument to show that we need to qualify conscious inessentialism because although “it is true that zombies who grew up in our midst might become glib in the use of our language, including our philosophical talk about consciousness [and other mentalistic concepts], a world of zombies could not originate these exact concepts.” (199)

Let’s take the argument step-by-step. Step-one is to concede the possibility of Turing-identical zombies growing up in our midst. So, for example, one might imagine a dramatically improved robotic version of NET-talk, a connectionist machine, that learns the language in its surround as humans do, and then uses our language, including terms like ‘believe’, ‘dream’, ‘see,’ and so on, just as we do.

But step-two involves denying that a zombie, or an isolated population of zombies, could “originate” our mentalistic vocabulary. There are many senses of “could,” so we need to be careful. Is the claim that it is metaphysically/logically impossible that the word “consciousness” and its suite could originate among creatures that are not conscious; or is it that it is nomically impossible in the sense that such origination violates the laws of nature that operate in our vicinity? Or is it simply a strong “implausibility” claim? For example, that it is highly unlikely that the origination of a mentalistic vocabulary would occur among an isolated group of zombies in the way it is highly unlikely that the Pope and Mother Teresa will have a

---

<sup>4</sup> This journal, 1, No. 2, Winter 1994, pp. 196-200.

picnic on the moon within the next month. It is metaphysically and logically possible that they will picnic, and Neil Armstrong and NASA proved that the laws of nature allow humans to go to the moon.

The point is that the sense of “could” matters a great deal. No one thinks that the existence of zombies who display no “mark of zombiehood” is likely in this actual world. It is just that the existence of zombies who are behaviourally indistinguishable from us appears to be metaphysically, logically, and nomically possible. And the reason these kinds of possibility matter is that they push us to sharpen our theories about the nature, function, and criteria of consciousness. If systems “just like us” could exist without consciousness, then why was this ingredient added? Does consciousness do something that couldn’t be done without it?—in addition, that is, to bringing experience into the world?

### 3. Zombie Earth

In order to determine the sense of possibility Moody is operating with when he concludes that conscious inessentialism requires “qualification” (199) we need to look more closely at the details of his thought experiment. Moody asks us to imagine an entire planet populated by zombies:

Suppose there is a world much like our own, except for one detail: the people of this world are insentient. They engage in complex behaviours very similar to ours, including speech, but these behaviours are not accompanied by conscious experience of any sort. (196)

We may think of this conveniently in terms of a Zombie Earth where each of us has a zombie doppelgänger.

Moody argues that there will indeed be behavioural differences between us and the inhabitants of Zombie Earth, differences that will manifest themselves at “the level of speech communities.” (197) Zombies, lacking the inner life that is the referent for our mentalistic terms, will not have concepts such as ‘dreaming’, ‘being in pain’, or ‘seeing’. This, Moody says, will reveal itself in the languages spoken on Zombie Earth, where terms for conscious phenomena will never be invented. The inhabitants of Zombie Earth won’t use the relevant mentalistic terms and thus will show “the mark of zombiehood” (199).

Moody suggests that if conscious inessentialism is true, then it is possible not just that an entire world like ours (except, of course, that it is inhabited by zombies) could exist, but moreover that such a world could evolve to be just like ours, but without any conscious life. Metaphysics and logic allow this possibility. Furthermore, the laws of nature, as understood so far, allow it.

Step-two, therefore, involves not merely imagining a Zombie Earth stipulated into existence, nor even imagining the possibility of connectionist zombies growing up among us (a possibility conceded in step-one), but imagining such a world developing according to the laws of evolution.

But step-three points out that this thought experiment will fail. Undetectable-evolved-zombies are impossible. They are unimaginable. This is because the zombie inhabitants of Zombie Earth would be distinguishable from us, because, lacking conscious lives, the zombies would never—indeed, could never—develop the mentalistic concepts and vocabulary that we have. According to Moody even though “the activities of talking about” such things

as dreams, or seeing “do not require consciousness, the emergence of those concepts in a language community does” (199). The zombies on Zombie Earth would display “the mark of zombiehood” in virtue of possessing no mentalistic vocabulary.

#### 4. How to Imagine Zombie Earth

Suppose this were true. Why does it matter? Moody thinks that “If conscious inessentialism is true, then it would presumably be impossible for us to tell whether visitors from another world are zombies. After all, if there is no necessary behavioural difference between them and us, as conscious inessentialism requires, there will be no identifiable mark of zombiehood.” (197)

The argument appears to be:

1. If conscious inessentialism is true, then it would not be possible to distinguish zombies from conscious creatures.
2. But it would be possible to tell that the inhabitants of Zombie Earth are zombies.
3. Therefore, conscious inessentialism is false, or in need of qualification.

The argument is valid, a clear case of the valid rule modus tollens, but it is unsound. Premise one is false. Conscious inessentialism says only that for any intelligent activity i, performed in any cognitive domain d, even if we do i with conscious accompaniments, i can in principle be done without these conscious accompaniments. The thesis says nothing about situations in

which the presence or absence of consciousness will or will not in fact be detectable.

Premise two is the claim that the zombies on Zombie Earth will be easy to detect as zombies. If they are as Moody paints them, lacking all mental terms, then this is probably true. But the issue at stake is a modal one. It is about possibility and necessity. If Moody thinks it is impossible in the strict sense—as opposed to merely implausible—for the occupants of Zombie Earth to be undetectable as such, he is simply wrong and premise two is also false.

This is easy to see. The occupants of Zombie Earth are smart. They are “informationally sensitive,” but they are not, being zombies, “experientially sensitive” (Flanagan 1992). However, they make speech sounds and record regularities they detect in their world using languages much like ours. According to Moody, this means that their mathematics and natural science “would likely be very similar to our own” (197).

But zombies will not originate speech sounds or written symbols for regularities that do not exist in their world. And so it is with conscious mental patterns. Zombie Earth is void of these. Since one doesn’t need terms for what doesn’t exist, “mentalist words” would never originate on Zombie Earth.

It is, to be sure, highly unlikely—implausible to the extreme—that mentalistic vocabulary would evolve among Moody’s zombies. But is it metaphysically, logically, or nomically impossible? No. There are (at least) two ways to think of how zombies might come to use mentalistic terms. One involves zombie cultural evolution; the other zombie species evolution.

Regarding cultural evolution, it might seem obvious that zombies, not being subjects of experience, would have no reason to develop terms that we might translate as ‘dreaming’ or ‘seeing.’ But consider this possibility: In the

beginning, informationally-sensitive zombies who detect (but don't experience) trees on Zombie Earth also notice that compatriot zombies sometimes bump into trees. Observation of this regularity leads to the invention of a warning locution such as "Watch out!" whose normal social function is to get zombies to turn around so that their photoreceptors receive the relevant message and dispose them to re-orient their motion. They start calling having one's photoreceptors oriented in the right direction "seeing." To be sure, it is seeing<sup>Z</sup>, the non-conscious zombie counterpart to our seeing. The point is that we wouldn't know that it was seeing<sup>Z</sup> and not seeing; and they would call it "seeing".

One could imagine a similar story for dreams. Zombies recharge during what they call "sleep." Most every night there are various sorts of irregularities in the recharging process. Suppose these involve continued activation of the zombies' speech centers.

As in the case of the trees, flora and fauna generally, and the movements of compatriots, they are able to detect the relevant regularities. In the morning they come to report the narrative sequence of speech center activation (yes, this is all "mindless"): "I saw a tree. Jane was in it. We had sex on a limb..." They come to call these reports "dreams." They are dreams<sup>Z</sup> not dreams, but the point is that such terms could originate among the zombies on Zombie Earth. The absence of such terms, or translatable equivalents, in zombie linguistic communities, could not serve as a "mark of zombiehood." It could not serve as the "mark of zombiehood" because such terms could be present after all!

Dream<sup>Z</sup> reports might even serve good zombie functions. What is activated in a zombie's speech center during sleep might helpfully clue in fellow zombies to expected behaviours, or to how well recharging went.

Activation of the speech centers might be a help or a hindrance to recharging—this too the zombies will discover<sup>Z</sup> and articulate.

With regard to zombie species evolution, Moody concedes, at least for purposes of his thought experiment, that evolution allows that Zombie Earth is a possibility. Here he appears to mean it is not only metaphysically and logically possible, but also nomically possible. The laws of nature as we know them in our vicinity allow that very intelligent, informationally-sensitive, but non-conscious creatures, could evolve. To the best of our knowledge, this is true.

But neo-Darwinian theory involves not just adaptations, but exaptations, free riders, and all manner of serendipitous developments. There is no metaphysical, logical, or nomic impossibility in the idea that Mother Nature selected an innate language of thought among zombies with lexical space for short-hand behavioural descriptions of the sort mentioned above (“seeing” for “turning photoreceptors in the-right-direction”; “dreaming” for “speech center activity during sleep”)—terms that seem to have mentalistic referents to us, but that do not.

## 5. Zombie Philosophy

Suppose it is conceded that the zombies on Zombie Earth could possibly originate a mentalistic vocabulary. There is still one move open to Moody. Even if we imagine that the zombies will utter sentences of the form “I see red” and “I dreamed that I visited my long lost cousin Jane and we did naughty things in a tree,” we will not be able to imagine the zombies philosophizing in the ways we do about seeing and dreaming. Zombies won’t be able to do philosophy of mind. It is this then, a lack of philosophical talent, that will mark them as zombies!

Indeed, in the course of developing his Zombie Earth scenario, Moody speculates about zombie philosophy and makes the claim that neither mental terms nor the philosophy of mind could originate among his zombies. He is wrong, as we have seen, about the first point. What about the second? Wrong again.

Moody suggests that zombie philosophers “would be especially puzzled<sup>Z</sup> by our human philosophical literature about dreaming.” (198) But zombie philosophers would not be puzzled<sup>Z</sup> at all, especially if their dreaming<sup>Z</sup> was as imagined above: speech center activation reported in the morning. It would turn out that they were wrong in taking us to be talking about dreams<sup>Z</sup> when really we were talking about dreams—but neither we nor they would ever be able to discover this from our respective behaviours alone. Will the zombies have dreams? No, but as we have suggested, they will use the word ‘dream’ to talk about their dreams<sup>Z</sup>, and they will ask the very same questions about dreams<sup>Z</sup> that we ask about dreams. They will wonder<sup>Z</sup>, for example, whether the things they call ‘dreams’ occur during sleep, or whether they are reports of speech center activation associated with waking up. They will notice the bizarre character of sleep speech center activation relative to awake speech and awake behaviours, and they will discern the relevant patterns. For example, many of the behaviours described by speech center activation while asleep are not the sort of things a ‘nice zombie’ would do in the light of day.

So it is with the inverted spectrum problem, which for the zombies would be the inverted colour judgment problem, but indistinguishable for them and for us from the normal inverted spectrum problem. Suppose that normal zombies, upon seeing light of a certain wavelength  $\underline{x}$  go into a state that is the disposition to say, “that object is green”, and then they act on that

disposition.<sup>5</sup> All that is necessary for an inverted colour judgment problem is that behavioural pathways get crossed twice. In our case (i.e., the usual inverted spectrum problem) one of the pathways is supposed to be the qualitative look of colour, and the other a speech act. Zombies could have an equivalent problem with two non-conscious inversions—no ‘internal seeing’ (198) is required. First, when seeing an object that reflects a wavelength  $\lambda$ , the inverted colour judgment zombie enters the state that, in normal zombies, is the disposition to say, “that object is red.” However, due to the second crossed wire, the inverted colour judgment zombie’s “that object is red” state actually causes it to utter, “that object is green.” Thus a double inversion can create a problem indistinguishable from the inverted spectrum problem.

We said that zombies are sometimes used to illustrate the problem of other minds; could zombies themselves have that philosophical concern<sup>Z</sup>? Yes, the zombies could consider<sup>Z</sup> a problem that they might call, oddly enough, “the problem of other minds.” How could this be?

Recall that the only difference between Zombie Earth and Earth is that its inhabitants are zombies. Importantly, there is no reason to think that the zombies believe<sup>Z</sup> (let alone believe) that they are zombies in our sense. In fact, by hypothesis they would call themselves “conscious.” Even after meeting us the zombies would not realize that they were zombies. They would be in exactly the same position with regards to us (and each other) that we are in with regards to them (and one another): they would have no overt behavioural evidence that we were any different than they. But all this is not

---

<sup>5</sup> We do not intend to suggest that colours be identified with wavelength, or even more complex surface reflectance data. We are merely constructing what the analogous problem would be for a zombie philosopher, who would have, by hypothesis, only external stimuli to go on.

because zombies could not make a contrast between “conscious creatures like themselves” and “zombies.”

Suppose one of those same zombie philosophers who wonders<sup>Z</sup> about dreams and colour inversion comes to ask a more basic question. The zombie philosopher would believe<sup>Z</sup> that she has ‘dreams’ and ‘sees’ objects. She might even know<sup>Z</sup> a bit about the brain states on which her ‘dreaming’ and ‘seeing’ supervene. In fact, it is the case that she really has dreams<sup>Z</sup> and sight<sup>Z</sup>. Part of what she means when she calls herself ‘conscious’ is that she has dreams<sup>Z</sup> and sight<sup>Z</sup>.

The zombie philosopher might wonder whether all of the other zombies are conscious<sup>Z</sup> in the way she is. Are their ‘dreams’ really, as she has discovered of her own, reports of speech center activity (and let us assume, visual center activity) that occurs during REM sleep; or are their so-called “dreams” something really weird like coming-to-say-things upon waking. Once the zombies have come to articulate this contrastive pattern, it is easy to suppose that they express the problem of wondering about other zombies in this form: Are our compatriots “conscious” like us or are they “zombies”? The zombie philosopher will wonder whether some of her compatriots are “zombies” whose brain states produce in them the “right” behaviours but who simply don’t go into anything like the same states they do. This much indicates how the “problem of other minds” could arise for the zombies.

## 6. The Hard Problem

The thesis of conscious inessentialism, the claim that intelligent activity without consciousness is possible -- across various senses of possibility -- has been defended against Moody’s Zombie Earth thought experiment. If it were true that a world of non-conscious creatures

functionally indistinguishable from us were logically impossible, then conscious inessentialism would be false or in need of qualification. But, according to the argument above, neither metaphysics and logic nor even the laws of nature stand in the way of this possibility. Zombies reign.

Now it is time to return to earth—both from Zombie Earth and the rarefied heights of modal logic and conceivability thought experiments. It is time to remind ourselves why philosophers need zombies, and why zombie thought experiments abound. As we said at the outset, there are several purposes served by zombie thought experiments: such thought experiments have relevance to claims made by functionalists, by advocates and opponents of the Turing-test, and of artificial intelligence, and more recently even, by those concerned about the prospects of artificial life. Thinking about the possibility of zombies sheds light on issues in the philosophy of language, especially, issues concerning the inscrutability of reference and the indeterminacy of translation, and in metaphysics, on issues of ontological relativity, for example.

In closing we focus on one issue whose difficulty is underestimated, but which Moody's Zombie Earth draws vivid attention to. This is the question, "Why are we conscious?"

The original discussion of conscious inessentialism leads off with this claim: "Consciousness did not have to evolve. It is conceivable that evolutionary processes could have worked to build creatures as efficient and intelligent as we are, even more efficient and intelligent, without these creatures being subjects of experience...However, from the fact that consciousness is inessential to highly evolved intelligent life, it does not follow that it is inessential to our particular type of intelligent life"<sup>6</sup>

---

<sup>6</sup> See Flanagan, 1991, p. 344; 1992, p. 129.

The point is that, skeptical worries to one side, Homo Sapiens are conscious. Assuming this is true, but that it is also true that there was no metaphysical, logical, or nomic necessity in making us so, why did Mother Nature settle on “being subjects of experience” as a good solution strategy for us, and quite possibly for numerous other mammals and other genera?

Some philosophers, e.g., David Chalmers, Joseph Levine, Colin McGinn, Thomas Nagel, think that the hardest problem is explaining how brain states could give rise to phenomenal states. This is indeed a hard problem. But this problem is certainly no more difficult than the problem of why and how it is that there came to be conscious creatures at all. Why did evolution result in creatures who were more than just informationally sensitive? There are, to the best of our knowledge, no good theories about this, and it is one of the reasons for pressing the conscious inessentialism worry.

Surely we jest, the reader might think. There must be good theories for why consciousness evolved. Well we have looked far and wide and no credible theories emerge. We can't respond to all comers here, but consider some recent contenders: The first tries to give an adaptationist account of the phenomenon of consciousness generally, the next two deal with particular kinds of consciousness—pain awareness and feelings of lust. The last appeals to consciousness as an explanation for certain complex mental capabilities.

The Stream of Consciousness. It is often said that the mind/brain must be a parallel-distributed-system, computing many different things at once. Consciousness, the serial one-thing-at-a-time stream, is a brilliant solution strategy for giving the “control center” of the whole system only the information it needs. There is so much the control center doesn't need to

know, why bother it with anything but the most important and highest quality information? The idea even has intuitive and theoretical appeal. After all, we know that at each moment very much is going on inside us of which we are entirely unaware. But our conscious stream seems to be just that, a smooth stream meandering along, not a chaotic jumble of parallel information processing.

Suppose this basic proposal is a good answer to problems like the computational bottleneck of a fully serial mind/brain, and to the problem of information overload. Still, it explains precisely nothing about why the information passed on to the (hypothesized) control center is conscious information as opposed to just information, of the sort for example that unconscious PDP computers utilize all the time.

Pain. Dennett (1990) writes that there “can be no doubt that having the alarm system of pain fibers and the associated tracts in the brain is an evolutionary boon, even if it means paying the price of having some alarms ring that we can’t do much about.” He then adds the immediate question: “But why do pains have to hurt so much? Why couldn’t it just be a loud bell in the mind’s ear, for instance?” (1990, 61).

Dennett is as sensitive to the problem of the adaptive role of consciousness as anyone. But like most everyone else, he doesn’t offer a good answer to his question, or to a variation on it: why couldn’t the system be constructed so that it detects things that are bad for it by an innate stock of, among other things, temperature sensitive sensors and blood-is-about-to-be-lost sensors that are hooked up to the right action paths.

Lust. The same sort of argument goes for certain pleasant experiences. Just after his above-quoted musings about pain Dennett writes: “And what, if anything, are the uses of anger, fear, and hatred (I take it the evolutionary utility of lust needs no defense.)” Why does Dennett think the evolutionary utility of lust is obvious? The reasoning is probably something like this: given that Mother Nature cares only about inclusive genetic fitness which requires reproduction (which does not for all species involve sexual reproduction, let alone fun), it is a good idea to create a powerful incentive to want to have sex with the right things. Lust creates the powerful incentive. On average, lust is directed towards enough of the right things to cause the right fitness enhancing behaviours in humans.

This much seems right. The trouble is that as Dennett points out before making the “obvious point” about lust: “There is almost nothing sexy (in human terms) about the sex life of flowers, oysters, and other simple forms of life.” (Dennett, 1990, 173). And Dennett points out that why there are conscious selves, and moreover lusty ones are “among the deepest problems in contemporary evolutionary theory” (1990, 172).

We need to reproduce. This is true. But why is the evolutionary utility of lust itself obvious? Olfaction, for example, figures importantly in the lives of many creatures that differ from humans in being sexually interested only sometimes. Male luna moths are sensitive to certain female moth pheromones. Sensing these gets them flying miles to the right location for mating. (Humans have been known to travel even further, and often with less encouragement). Why think, even supposing that luna moth mating behaviour is regulated by olfactory cues, that any advantage whatsoever is conferred by experienced odours that arouse lust over sensors

that unconsciously pick-up the airborne chemical cues indicating that it is time to mate? There is at present no good answer to this question.

Learning and Plasticity. One might acknowledge this possibility, for example, that our pain and sex lives might be regulated by unconscious information pick-up, but claim that the organism needs to learn. If there are zombies in our midst, why not pheromones from things one ought not to want to mate with, or hot things that one must touch -- and toss away -- to save one's group even though the temperature readings are just like "Move and Run" signals?

It is this that consciousness is for. It confers, like nothing else could, plasticity. The wired-up sex or harm detection systems might contain a helpful stock of sex-initiating or harm-avoiding-routines, but it is "feeling sexy" or "feeling pain" that will be key to a creature's learning all the idiosyncratic facts about what to mate with or what is unsafe in its unique ecological niche. Innate responses to basic evolutionarily advantageous or disadvantageous things might get us to mate or avoid standard bad things, but they wouldn't get us to learn about the contingent features of our environment on which rests our ultimate success.

This argument won't work. Plasticity, learning, and the like need not be, indeed in our own case they often are not, conscious. Computers show that learning and plasticity without awareness is possible, as do humans.

It may well be, we suspect it is the case, as a matter of contingent fact, that for human beings consciousness does facilitate learning, that it does radically influence behaviour, that it does influence higher-level thought processes. The above stories are not stories about how we work, they are stories about how we might have worked. Recognition that consciousness did not have to be highlights the pressing questions of why it exists, why it

evolved, whether there were competing zombie hominids who lost out in the struggle to survive, and if there were not why not?

Telling a convincing story about the adaptive advantage of consciousness is very hard. The upshot is that there exist no good stories for why consciousness evolved in this actual world. There are as yet no credible story about why subjects of experience emerged, why they might have won—or should have been expected to win—an evolutionary battle against very intelligent zombie-like information sensitive organisms has yet been told. At least this has not been done in a way that provides a respectable theory for why subjects of experience gained hold in this actual world—for why we are not zombies.

Our lack of such a theory is one of the reasons why we need zombies. They keep things interesting in the philosophy of mind. We might have been zombies. We are not. But it is notoriously difficult to explain why we are not.

This problem, the problem of the evolution of consciousness, is a problem for biologists, cognitive scientists, and philosophers of mind with interest and expertise in evolutionary theory. We need theoretical cooperation from such people plus comparative psychologists, ethnologists, paleontologists, zoologists, and neuroscientists, among others. It is hard to see how the artificial intelligentsia will contribute to the next stage of work. They have helped enormously already, setting the agenda by creating zombies in our midst. They have also shown us that the problems of intelligence and consciousness divide to a point. To solve the two hard problems: how consciousness supervenes on states of matter and why it evolved in the first place requires close study of systems that are conscious, biological creatures that have evolved to be subjects of experience. Us.