

SYMPOSIUM ON 'CONVERSATIONS WITH ZOMBIES'

Journal of Consciousness Studies

*controversies in science
& the humanities*

an international
multi-disciplinary journal

Volume 2, No.4, 1995

***Rodney Cotterill:
'On the unity of conscious experience'***

Zombie Earth: a symposium

Daniel Dennett Jaron Lanier

Owen Flanagan Thomas Polger

Selmer Bringsjord Charles Tart . .

***and other commentators on Todd Moody's
'Conversations with zombies'***

THE UNIMAGINED PREPOSTEROUSNESS OF ZOMBIES

*Daniel C. Dennett, Tufts University, Center for Cognitive Studies
11 Miner Hall, Medford, Massachusetts 02155-7059, USA*

Abstract: Knock-down refutations are rare in philosophy, and unambiguous self-refutations are even rarer. For obvious reasons, but sometimes we get lucky. Sometimes philosophers clutch an insupportable hypothesis to their bosoms and run headlong over the cliff edge. Then, like cartoon characters, they hang there in mid-air, until they notice what they have done and gravity takes over. Just such a boon is the philosophers' concept of a zombie, a strangely attractive notion that sums up, in one leaden lump, almost everything that I think is wrong with current thinking about consciousness. Philosophers ought to have dropped the zombie like a hot potato, but since they persist in their embrace, this gives me a golden opportunity to focus attention on the most seductive error in current thinking.

Todd Moody's (1994) essay on zombies, and Owen Flanagan and Thomas Polger's commentary on it, vividly illustrate a point I have made before, but now want to drive home: when philosophers claim that zombies are conceivable, they invariably underestimate the task of conception (or imagination), and end up imagining something that violates their own definition. This conceals from them the fact that the philosophical concept of a zombie is sillier than they have noticed. Or to put the same point positively, the fact that they take zombies seriously can be used to show just how easy it is to underestimate the power of the 'behaviourism' they oppose. Again and again in Moody's essay, he imagines scenarios to which he is not entitled. If, *ex hypothesi*, zombies are behaviourally indistinguishable from us normal folk, then they are really behaviourally indistinguishable! They say just what we say, they understand what they say (or, not to beg any questions, they understand² what they say), they believe² what we believe, right down to having beliefs² that perfectly mirror all our beliefs about inverted spectra, 'qualia' and every other possible topic of human reflection and conversation. Flanagan and Polger point out several of Moody's imaginative lapses on these matters in careful detail, so I needn't belabour them. In any case, they follow trivially from the philosophical concept of a zombie.

Flanagan and Polger also fall in the very same trap, however. For instance, they say it is 'highly unlikely — implausible to the extreme — that mentalist vocabulary would evolve among Moody's zombies. But is it metaphysically, logically, or nomically impossible? No.' Here getting it half right is getting it all wrong. It is *not at all* unlikely or implausible that mentalistic vocabulary could evolve among zombies. That must be conceded as part of the concession that zombies are 'behavioural' twins of conscious beings; if it is likely that we conscious folks would develop mentalistic vocabulary, then it must be exactly as likely that zombies do. It is just such lapses as this one by Flanagan and Polger that feed the persistent mis-imagination of zombies and make them appear less preposterous than they are.

Some zombies *do* have an 'inner life'. In Dennett (1991), I introduced the category of a *zimbo*, by definition a zombie equipped for higher-order reflective informational states (e.g., beliefs² about its other beliefs² and its other zombic states). This was a strategic move on my part, I hasten to add. Its point was to make a distinction within the imaginary category of zombies that would have to be granted by believers in zombies, and that could do all the work they imputed to consciousness, thereby showing either that their concept was subtly self-contradictory, since some zombies — zimboes — were conscious after

all, or that their concept of consciousness was not tied to anything familiar and hence amounted to an illicit contrast: consciousness as a 'player to be named later' or an undeclared wild-card. As I pointed out when I introduced the term, zombies behaviourally indistinguishable from us are zimboes, capable of all the higher-order reflections we are capable of, because they are competent, *ex hypothesi*, to execute all the behaviours that, when we perform them, manifestly depend on our higher-order reflections. Only zimboes could pass a demanding Turing Test, for instance, since the judge can ask as many questions as you like about what it was like answering the previous question, what it is like thinking about how to answer this question, and so forth. Zimboes think² they are conscious, think² they have qualia, think² they suffer pains — they are just 'wrong' (according to this lamentable tradition), in ways that neither they nor we could ever discover!

According to Flanagan and Polger, there is still a difference between the inner lives of zombies and ours: theirs are merely 'informationally sensitive' while ours are also 'experientially sensitive'. This contrast, drawn from Flanagan (1992) is ill-conceived, so far as I can see. It apparently arises out of a self-induced illusion, a fallacy of subtraction. On this outing, it arises in Flanagan and Polger's brisk dismissal of my accounts of pain and lust. They concede that there is an uncontroversial evolutionary explanation of why there should be a peremptory (well-nigh irresistible) damage-alarm system, or a clarion call system in the case of sex, but, they ask, what about the evolutionary explanation of the conscious feelings themselves? They see me as failing to address (or address well) my own further questions about why the pains should hurt so much, or why (in our species) lust should be so lusty. I had better spell out my view in a little more detail.

In creatures as cognitively complex as us (with our roughly inexhaustible capacity for meta-reflections and higher-order competitions between policies, meta-policies, etc.), the 'blood-is-about-to-be-lost sensors' and their kin cannot simply be 'hooked up to the right action paths' as Flanagan and Polger put it. These ancestral hookups, which serve quite well in simpler creatures, do exist in us as part of our innate wiring, but they can be overridden by a sufficiently potent coalition of competing drafts (in my Multiple Drafts Model). So it is that in us, there is an *informationally sensitive* tug-of-war, which leaves many traces, provokes many reflections (about the detailed awfulness of it all), permits many protocols (and poignant complaints), adjusts many behavioural propensities (such as desires to reform one's ways, or seek revenge), lays down many ('false') memories about the way the pain felt, etc., etc. Notice that 'experiential sensitivity' (whatever that might be) never comes into my account — the 'conscious feelings themselves' get distributed around into the powers and dispositional effects of all the extra informational machinery — machinery lacking in the case of simple organisms. My account is entirely in terms of complications that naturally arise in the competitive interaction between elements in the Multiple Drafts Model. It would serve exactly as well as an account of the pain and lust systems in zimboes, who are so complex in their internal cognitive architecture that whenever there is a strong signal in either the pain or the lust circuitry, all these 'merely informational' effects, and the effects of those effects, etc.) are engendered. That's why zimboes, too, wonder² why sex is so sexy for them (but not for simpler zombies, such as insects) and why their pains have to 'hurt'.

If you deny that zimboes would wonder such wonders, you contradict the definition of a zombie. Hark back to a point I made in *Brainstorms* (1978) about what I called the beta-manifold of beliefs arising from having mental images. The game is over as soon as the zombie-defender grants — as the zombie-defender must on pain of self-contradiction — that there is no human belief (however much concerned with 'phenomenal' topics) that

does not have a zombic functional counterpart — a mere belief² — which is behaviourally indistinguishable from that belief.

Notice what has happened. The contrast Flanagan and Polger would draw between zombies and us in regard to pain and sex, I draw between simple zombies and fancier zombies — zimboes. And I have offered a sketch of an evolutionary account of why this contrast would exist. (It is evolutionary in the extended sense that applies to systems that evolve within individual organisms during their lifetimes.) This responds directly to Flanagan and Polger's demand for an explanation from me of why *our* pains should hurt and why sex should strike *us* as so sexy. We are not just simple zombies, but zimboes, and zimboes, unlike insects and even most mammals, *do* find sex sexy, and *do* wonder about why pain has to hurt. Flanagan and Polger overlook this fact about zimboes, and hence mis-locate the contrast. 'Why think,' they ask, '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?' But who says luna moths do experience lust? I see no reason to believe they do. If luna moths are zombies, they are not zimboes; they are simpler zombies. And indeed, to respond to Flanagan and Polger's question, there is no evolutionary reason at hand to suppose that their sex-drive is complicated by storms of lust. The experience of lust is the fate of fancier creatures — zimboes like us.

This mislocation of the issue by Flanagan and Polger gives illicit support to their claim that 'experiential sensitivity' differs from mere 'informational sensitivity' by encouraging (however unintendedly) the following fallacious train of thought (I have tested this on my students): zombies would pull their hands off hot stoves, and breed like luna moths, but they wouldn't be upset by memories or anticipations of pain, and they wouldn't be apt to engage in sexual fantasies. No. While all this might be true of simple zombies, zimboes would be exactly as engrossed by sexual fantasies as we are, and exactly as unwilling to engage in behaviours they anticipate to be painful. If you imagine them otherwise, you have just not imagined zombies correctly.

Flanagan and Polger compound this mistake when they go on to ask what the adaptive advantage of consciousness (as contrasted with mere 'informational sensitivity') would be. It can't be 'learning and plasticity', they argue, since learning and plasticity can go on without consciousness. It can go on in zimboes, for instance. I expect that they might concede that my story does succeed in explaining (in a sketchy and speculative way) why zimboes might evolve from simpler zombies, and what particular advantages (e.g. in learning and plasticity) their capacities for higher-order reflection might have. But they don't see this as even the beginnings of an evolutionary account of consciousness, since by their lights what zimboes have but simpler zombies lack is not consciousness at all, not 'experiential sensitivity'. What zimboes have, however, is the very sort of thing they were pointing to when they demanded from me an account of hurting pains and sexy feelings.

The question of adaptive advantage, however, is ill-posed in the first place. If consciousness is (as I argue) not a single wonderful separable thing ('experiential sensitivity') but a huge complex of many different informational capacities that individually arise for a wide variety of reasons, there is no reason to suppose that 'it' is something that stands in need of its own separable status as fitness-enhancing. It is not a separate organ or a separate medium or a separate talent.

To see the fallacy, consider the parallel question about what the adaptive advantage of *health* is. Consider 'health inessentialism': for any bodily activity *b*, performed in any

domain *d*, even if we need to be healthy to engage in it (e.g. pole vaulting, swimming the English Channel, climbing Mount Everest), it could in principle be engaged in by something that wasn't healthy at all. So what is health *for*? Such a mystery! But the mystery would arise only for someone who made the mistake of supposing that health was some *additional* thing that could be added or subtracted to the proper workings of all the parts. In the case of health we are not apt to make such a simple mistake, but there is a tradition of supposing just this in the case of consciousness. Supposing that by an act of stipulative imagination you can remove consciousness while leaving all cognitive systems intact — a quite standard but entirely bogus feat of imagination — is like supposing that by an act of stipulative imagination, you can remove health while leaving all bodily functions and powers intact. If you think you can imagine this, it's only because you are confusedly imagining some health-module that might or might not be present in a body. Health isn't that sort of thing, and neither is consciousness.

All I can do at this point is to reiterate my plea: consider the suggestion, once again, that when you've given an evolutionary account of the talents of zimboes, you've answered all the real questions about consciousness *because the putative contrast between zombies and conscious beings is illusory*. I know that many philosophers are sure that it is not illusory. I know that they are sure that *they* don't make such mistakes of imagination when they claim to conceive of zombies. Maybe they don't. But Moody does, as Flanagan and Polger show, and then they themselves do, as I have just shown. These two essays continue a tradition unbroken in my experience. That is, I have never seen an argument in support of the zombie distinction that didn't make a mistake of the imagination of this sort. This prevailing wind of mis-imagination undoubtedly provides ill-gotten support for the widespread conviction among the populace (reported, for instance, by David Chalmers) that there is an important difference between zombies and conscious beings. That is, those who believe in the distinction ought to grant me that even professional philosophers sometimes do fail to appreciate that these are *bad* reasons for believing in the distinction. The question I pose to them is whether there are any *good* reasons for believing in the distinction. My conviction is that the philosophical tradition of zombies would die overnight if philosophers ceased to mis-imagine them, but of course I cannot prove it *a priori*. We will just have to wait for some philosopher to write an essay in defence of zombies that doesn't commit any such misdirections, and see what happens.

To make sure the challenge is clear, let me review the burden and its attendant requirements. One must show that there is a difference between conscious beings and zombies, *and* one must show that one's demonstration of this difference doesn't depend on underestimating in the well-nigh standard way the powers of zombies. Here's a handy way of checking one's exercises of imagination: demonstrate that a parallel difference does *not* exist between zimboes and less fancy zombies. One may in this way ensure that one hasn't simply underestimated the power of zombies by imagining some crude non-zimbo zombie, rather than a zombie with all the 'informational sensitivity' of us human beings.

In a thoughtful essay Joseph Levine (1995) deplores the ridicule to which I have subjected believers in zombies. This is a matter of some urgency for philosophers, since, as I say in 'Get Real' (Dennett, 1995 — my reply to Levine and others), it is an embarrassment to our discipline that what is widely regarded among philosophers as a major theoretical controversy should come down to whether or not zombies (philosophers' zombies) are possible/conceivable. I deplore the bad image this gives to philosophy just as much as Levine does, but for the life of me, I can't see why a belief in

zombies isn't simply ridiculous, and I'm going to go on comparing zombies to epiphenomenal gremlins and other such *preposter*a until some philosopher mounts a proper defence, showing that the belief in the possibility of zombies is somehow better supported than these other cases.

Again, let me clarify the challenge. It seems to me that postulating zombies is exactly as silly as postulating epiphenomenal gremlins, and so when a philosopher does it, I blush for the profession. Show me, please, why the zombie hypothesis deserves to be taken seriously, and I will apologize handsomely for having ridiculed those who think so. But I want to see an argument, and not just the nudges and mumbles I have provoked by this question in the past. Tradition doesn't cut it. 'If you got to ask, you ain't never gonna get to know' doesn't cut it. 'Everybody else believes in them' doesn't cut it. Calling me an instrumentalist or an operationalist or a behaviourist — as if these were obviously terms of censure — doesn't cut it. If the philosophical concept of zombies is so important, so useful, some philosopher ought to be able to say why in non-question-begging terms. I'll be curious to see if anybody can mount such a defence, but I won't be holding my breath.

References

- Dennett, D.C. (1978), *Brainstorms* (Montgomery, VT: Bradford Books).
- Dennett, D.C. (1991), *Consciousness Explained* (Boston, MA: Little, Brown and Company).
- Dennett, D.C. (1995), 'Get real', *Philosophical Topics*, 22, pp. 505–68.
- Flanagan, O. (1992), *Consciousness Reconsidered* (Cambridge, MA: MIT Press).
- Levine, J. (1995), 'Out of the closet: a qualophile confronts qualophobia', *Philosophical Topics*, 22, pp. 107–26.
- Moody, T.C. (1994), 'Conversations with zombies', *Journal of Consc. Studies*, 1 (2), pp. 196–200.