On the Conceivability of Zombies
(Chalmers v. Dennett)

Allin Cottrell*

Second draft, October 1996

"Any story about physical process applies equally to me and to my zombie twin. It follows that nothing in that story says why, in my case, consciousness arises. . . . The very fact that it is logically possible that the physical facts could be the same while the facts about consciousness are different shows us that. . . there is an *explanatory gap* between the physical level and conscious experience."

— David Chalmers (1996, p. 107)

"But just try to keep hold of this idea in the midst of your ordinary intercourse with others, in the street, say! Say to yourself, for example: 'The children over there are mere automata; all their liveliness is mere automatism.' And you will either find these words becoming quite meaningless; or you will produce in yourself some kind of uncanny feeling, or something of the sort."

— Ludwig Wittgenstein (1953, p. 126)

1 Introduction

A zombie—in the philosophers' rather than the Hollywood sense of the term—is a being that is physically (and hence also functionally) exactly like an ordinary human being, but which has no conscious experiences. The notion of zombies in this sense does a great deal of work in David Chalmers's recent book. (As a glance at the index confirms: "zombie" is one of the most heavily indexed terms in the

*Department of Economics, Wake Forest University; cottrell@wfu.edu

book.) Chalmers's case for the "logical possibility" (or conceivability) of zombies is intended to dramatize and reinforce the basic intuition that informs his whole argument, namely that conscious states, or qualia, are entirely resistant to any sort of physical or functional analysis. Zombies also provide a useful lever in the opposite direction: if one could make a persuasive case for the *in*conceivability of zombies, this would undercut the notion of an "explanatory gap", leaving the way open for a functional analysis of consciousness of the sort offered by Dennett (1991). As things stand, Chalmers reckons that the zombie argument rules Dennett's account (along with any others along similar lines) out of court; they may be very interesting and all that, but they simply evade the hard problem of consciousness, since we can imagine all the sort of processing that Dennett describes going on in a being that has no qualia, no subjective consciousness.

Well, can we? I shall argue that we cannot. But, lest the following discussion appear excessively frivolous, it may be worth prefacing the main discussion with an amplification, or at least a restatement, of what is at stake. The "real" issue concerns the status of qualia, that is, the subjective sensory states into which we are thrown when (say) looking at a yellow leaf, hearing a musical chord, sniffing a camembert, or running our fingers over a piece of sandpaper. Is it possible to provide a satisfactory account of such states using only the resources of a materialist functionalism? Or is it the case—as it has seemed to many, and as it seems to Chalmers—that once we have said all there is to say about the physical basis of, and the functional role of, such states, there remains an uneliminable residue: the brute qualitative matter of "what it is like" to sniff the camembert? Since it is extraordinarily hard to tackle this question head-on we seek the leverage afforded by the notion of the philosophers' zombie, the point being that if we have a coherent intuition to the effect that there is indeed such a residue, then we ought to be able to conceive of the zombie. Just subtract the residue while leaving all the physical/functional stuff in place. Conversely, if it transpires that the notion of the philosopher's zombie breaks down under stress, this would seem to indicate that the intuition of the ineliminable residue is itself problematic.

2 Assessing conceivability: some exercises

Can we conceive of a unicorn? Leaving aside any magical properties that unicorns might be conceived as possessing, the task is not very difficult. The concept of a unicorn is the conjunction of two perfectly ordinary concepts (each having a real referent), namely the concept of a white horse and the concept of a single, straight horn. We conjoin the two by imagining the horn growing from the white horse's forehead. We can "verify" that this conjunction "works" by drawing a picture of a unicorn, or making a three-dimensional model of one. Given this, the burden of proof surely lies with anyone wishing to claim that unicorns are not properly conceivable: Tell us where the hidden contradiction, incoherency or paradox lies, since none is at all apparent.

Now what about a person who behaves perfectly normally, but who turns out to have nothing but sawdust inside his skull? This is much more of a stretch, but we can still make a definite argument for conceivability. The components of the imagined being are in themselves unproblematic. Easy enough to imagine a normally-behaving person; easy enough to imagine a skull filled with sawdust and bereft of a brain. We have trouble taking seriously the idea that these components could ever be combined in practice, of course, and this makes the brainless wonder rather different from the unicorn. We can imagine that unicorns might actually have been, had biological evolution on this planet taken a few different turnings, but the notion that there could be a person who behaves normally without the benefit of a brain flies in the face of everything we think we know about the material conditions for intelligence. (In Chalmers's terminology, unicorns are "naturally possible" while brainless wonders are not.) All the same, we can *imagine* this, in some detail if need be. A drawing or model would not work as verification of consistency in this case (it could convey the sawdust-filled skull, but not the normal behaviour), but we could make a movie about it. (X seems perfectly normal; one day he develops the symptoms of a minor neurological complaint; his brain scan is anomalously blank; X goes in for exploratory surgery; we witness the surgeon's amazement as he cuts his way into the sawdust-filled cavity. . .) There are some loose ends to be tidied up: for instance, if the skull contains no brain, what happens to the blood vessels that normally supply the brain? That is, we need to think through the question: What exactly are we supposing we would see if we opened sawdust man's skull? But given a little time and ingenuity, we could presumably figure out answers to such questions. In this case also, then, the burden would seem to lie with the person who claims the brainless wonder is not really conceivable. Granting that this imagining is not coherent with our picture of how the world works, we can nonetheless ask, wherein lies its *internal* incoherence?

3 Zombiehood

Are we ready to tackle the zombie? This concept is again formed via the conjunction of two elements. The first is the same as the first element in the imagining of sawdust man, namely a human being who behaves perfectly normally. No problem there. But what exactly is the second?

At a first pass, it is the notion of a being lacking any "internal life". Chalmers on several occasions uses the phrase "all dark inside" to capture this idea. This phrase is problematic for two reasons. First, if taken literally, it draws no contrast with an ordinary person. We are all "dark inside", since there is no luminous counterpart inside our heads to the perception of a blue sky or the yellow of an autumn leaf. But of course the phrase is intended metaphorically. Very well, but "darkness" is itself a *quale*—that corresponding to the absence of retinal stimulation. There is something "it is like" to look into a coal cellar. "Inside" the zombie there is neither subjective light nor subjective darkness, but mere nullity. Insofar as darkness is easier to imagine than nullity, the "dark inside" metaphor steals a free ride, and should be disbarred. So how else might one concretize that which we are supposed to be conceiving? One might, for instance, take an object that is paradigmatic of subjective nullity—a rock, for instance. A zombie, then, is like a person on the outside, but just like a rock on the "inside".

Now we have two terms to work with OK, but the problem is that it's not clear how we are supposed to stick them together. Indeed, the more one thinks about this, I submit, the more obscure it becomes, exactly what one is supposed to be imagining. Note that the task is not conceiving what it would be like to *be* a zombie—in the way one might try to imagine what it would be like to be a mouse, or the President of the United States, or a person blind from birth—since by stipulation there is *nothing* "it is like" to be a zombie. Neither, however, is the task simply conceiving of a zombie "from the outside". That is, in a sense, too easy, since from the outside a zombie is, by construction, indistinguishable from a normal person. (Nonetheless, Dennett has pointed out (e.g., 1995a) that people who claim to be able to conceive of zombies quite often manifestly fall down at the task: they let their supposition that there is nothing going on "inside" the zombie bleed over into their statements about the sort of behaviour the zombie is likely to exhibit, thus contradicting their own stipulation that zombies are, so far as behaviour or functionality is concerned, perfectly normal.) The task is somehow to hold in our heads simultaneously the notion of normal human behaviour, on the one hand, and internal nullity, on the other. But note that *conceiving* of this, properly so called, must be more than simply maintaining the *verbal formula*, "normal behaviour, nothing inside" with rigid consistency (as Chalmers manages—he does not fall into the obvious sort of inconsistency that Dennett diagnoses in many other writers).

In the first two cases considered above, we concluded that the burden of proof would rest with anyone claiming that the putative conception (unicorn, sawdust man) was not properly conceivable. One reason for this conclusion was that in those cases it was possible to construct an external "verification" of the consistency of the conception (the model unicorn, the movie about sawdust man), which, if not conclusive, at least establishes a prima facie case for conceivability.[1] But note that no such verification is possible, even in principle, in the case of the zombie. We could make a movie "about a zombie" alright, but if the job were done properly, nothing about the zombie himself (itself?) could so much as hint at his

[1]Not conclusive, for this sort of reason. Suppose the model unicorn is made of wood. Then a skeptic might say, "OK, you can combine a horse and a horn in wood alright, but your conception is not of a wooden being, and I will explain why you can't really conceive of this conjunction in the proper materials. . . ".

zombiehood. We could represent other people as worrying about the possibility that X was a zombie, but again, if the job were done properly, these worries would have to be represented as entirely unmotivated. Glassy-eyed stares, or reluctance on the part of the zombie to talk about his feelings—or at any rate, more of these features that we find in ordinary humans—would simply be cheating, by virtue of the "normal behaviour" stipulation.

Given this, it seems that a neutral umpire of the zombie issue might well take a more even-handed approach to the burden of proof. On the one hand, "X, will you please give us your reasons for thinking that zombies are not conceivable", but on the other, "Y, will you please keep talking about these zombies—give us a fuller idea of what you're on about, so we can see if the conception seems to makes sense."

4 Intermediate sensory blanking

If the brute conjunction of normal behaviour and nothing inside presents, so to speak, a glassy-smooth surface on which the imagination can find no toehold, perhaps we can approach the issue by dropping back a bit from fully fledged zombiehood. The zombie is subjectively blank in all sensory modalities, and sustains no (subjectively registered) inner conversation or "stream of consciousness" of any kind. Yet he behaves normally in all respects, including his verbal *reporting* of subjective perceptions and inner conversation. If *this* is conceivable, then presumably the following ought to be conceivable just as readily, if not more so: Somebody who has the full range of normal human functionality but who is subjectively blank with regard to vision (only). That is, he "sees" perfectly well in the objective, information-processing sense, but he has no visual phenomenology like we do. This person has what Dennett calls "super-blindsight". And, besides granting this person a subjective, inner life like ours in all but the visual modality, let's make one further amendment to the standard rules regarding zombiehood—in addition he has one (just one) behavioural abnormality: he sometimes talks about this funny vision of his. That is, we do not insist that his *verbal reportage* of what's going on visually, for him, is in every respect the same as that of an ordinary person—though we do of course insist that he can walk over rough ground, play hunt the thimble, discuss paintings, read the newspaper, and all the rest of it, with the best of us. We might also imagine that this person once enjoyed fully normal sight, including the subjective phenomenal accompaniment, but that he has now lost the latter.

Hic rhodus, hic salta. It is hard to see how a defender of the conceivability of zombies could refuse the challenge: Please imagine how a conversation with such a person, on the subject of sight, might go. Please use this specific opportunity to flesh out the verbal formula, "normal functionality, nothing going on inside", and thereby to establish its coherence (or expose its incoherence). Once again, if fully-fledged zombiehood is conceivable, this ought to be relatively easy. Let's try. In the mode of various samples given by Dennett (1994, 1995b), our super-blindsighter might for instance say

> It seems to me that outside my window just now the autumn leaves—lemon-yellow, orange and russet—are scintillating in the wind. The grass, on which the afternoon sun is casting long shadows, also shows the subtle browning of autumn. Nearer at hand, the sunlight is painting the bricks of the wall outside my office in sharply defined siennas and ochres, so that I'm reminded of the little bricks of watercolor in the paintbox I had as a child. Behind the trees, the sun is glancing off some parked cars, making bright highlights and giving the appearance that the tail lights on one of the cars are lit. It seems to me that the falling of the leaves is more advanced in the big tree that is further from my window, revealing a cascading tracery of branches. But mind you, I can't actually *see* any of this the way I used to—it comes to me in a glance, as before, but I'm simply "registering" it all.

The question is whether we really reckon we are able to make sense of this speech. Or, to put it differently, whether we can imagine crediting the last sentence, rather than concluding that the person making the speech was subject to some strange delusion. Dennett urges the intuition that we would never be inclined to credit it, and I find myself in agreement. I just don't seem able to form a conception of *this* sort of information's being acquired, via a few moments' gaze out of the window, *without* there being "something it is like" to acquire it, without there being an impression of subjective phenomenology. (Although it is easy enough to conceive of, say, the photocell at the grocery door registering my presence without any impression of subjective phenomenology being thereby generated.)

The problem may equally well be put in the second person. You wake up one morning, open your eyes, and... what do you notice first? That the sun is streaming in the window, that your alarm clock says 7:30, and that your partner is already getting dressed on the other side of the room—or that, despite registering all this in a moment, you can't actually *see* anything? Or try a more radical variant, one that takes us a step back towards full zombiehood. You wake up one morning to find that *all* of your sensory modalities have blanked out subjectively (though you're still getting all the information alright). Again, what do you notice first: the sunlight, the sounds of birdsong and of traffic in the distance, the smell of coffee brewing—or the total absence of sensory qualia?

The complaint might be made here, that I am loading the question (towards the conclusion that the description of your putative state is incoherent, which indeed seems to me to be the case) by talking of the *sound* of the traffic and the *smell* of coffee, in the absence of qualia. Well, put those terms in quotes if you like, but notice that, in some shape or form, they have a right to feature in the description. By assumption, you are getting *all* the information you usually get, via the usual physical sensory channels, along with all the usual behavioural dispositions. You are immediately and confidently aware that there's coffee brewing, and coffee of a particular sort (high roast); we're not talking about "hunches", or better than chance performance on forced guessing about what's going on in the kitchen. And with the sound, it's not a matter of a degraded perception of "traffic noise yes/no". You can distinguish that big truck accelerating right now over the background rush of tyres and engine noise—even though you can't actually *hear* anything.

Are you having any difficulty imagining this? Let's probe further. This "smell" of (chemical registration of the presence of) coffee: does it give you any pleasure?[2]

[2] I cannot get into the matter here, but Damasio (1994) and Dennett (1996) offer interesting accounts of the complex evolutionary basis of the affective dimension to our sensory states.

By virtue of the "normal behaviour" postulate, it creates a disposition to sniff deeply and say "Ahh". Does this seem natural to you, or does it seem a strange and pointless piece of behaviour, a sort of tic? If it seems natural, how could that be? Have you "failed to notice" that your registration of the presence of coffee no longer has the qualitative dimension it used to have? And what about the truck "noises" (i.e., your auditory registration of the sound waves generated by the trucks in the vicinity)? Do you still find them annoying? You still issue your customary verbal complaints, but do these seem like a senseless rigmarole or not? Once again, if your complaints are still heartfelt, perhaps this is because you somehow fail to notice that the detailed auditory information you're receiving is no longer accompanied by any genuine, subjectively nasty qualia? Are you really able to make any sense of the last suggestion?

5 Chalmers and "dancing qualia"

The last possibility canvassed above—namely, that one might fail to notice that one's sensory equipment had just flipped between "qualia-included" mode and "plain information" mode—is one that Chalmers raises in the context of his "dancing qualia" thought experiment (1996, ch. 7), *as a necessary consequence of the possibility of zombies*.[3]

A little background is required. Chalmers is certainly not arguing that such a scenario is *plausible*. On the contrary, he says that its extreme implausibility provides support for his principle of "organizational invariance" (same functional organization, same subjective sensory qualia, as a matter of natural necessity). On

[3]Chalmers devotes most of his discussion of "dancing qualia" to the case of *inversion* of qualia, but he notes (p. 270) that the argument can equally well be run in terms of presence/absence of qualia. Since the dancing qualia argument is one of the most celebrated features of Chalmers's book, and is presented clearly and concisely by its author, I shall not rehearse it here. Briefly, Chalmers shows that if an *exact* functional isomorph of yourself, lacking any qualia, is possible, then—given some relatively uncontroversial auxiliary assumptions that are needed to get a "partial replacement" scenario going, and given a materialist-functionalist analysis of psychological states such as belief—it follows that there must be circumstances under which you would fail to notice the disappearance and reappearance of your qualia.

the other hand, his argument requires that we be able to form a coherent conception of this scenario. Let me try to clarify. Chalmers is walking a tightrope. In chapter 3 of his book he is at pains to defend the idea that zombies—functional duplicates of normal people who nonetheless lack any qualia—are perfectly conceivable (logically possible). As we have seen, this is an integral part of his argument; it licenses the dismissal of rival theories of consciousness such as Dennett's. But in chapter 7 Chalmers is equally at pains to deny that the notion of a zombie is at all plausible; this claim forms an integral part of his own positive theory (dualist functionalism). Is there any contradiction here? Well, there's no doubt that in some cases we can draw a clear line between questions of conceivability and questions of credibility. Matters governed by well-defined probability laws perhaps provide the clearest cases. Is it *conceivable* that an ice-cube could form spontaneously in the middle of my hot cup of coffee? This possibility is not ruled out by the theory of statistical mechanics, but it is assigned a vanishingly small probability and I would be willing to bet anything you like against its occurrence. There would be nothing inherently suspect about an argument which capitalized on both the in-principle conceivability and the practical incredibility of the spontaneous ice-cube. But can the same be said of zombies? Here (obviously) there is no question of any definite probabilistic reasoning. Really, it is all a matter of "what we can imagine" (or think we can imagine), and this is notoriously slippery.

From this perspective, there is something rather tendentious about Chalmers's separation of the issues in chapters 3 and 7. When the concern is (to put it crudely) the trashing of ordinary materialist functionalism, he asserts the conceivability of zombies in quite facile terms, leaving the impression that one would have to be strangely lacking in imagination to deny this idea. But once ordinary functionalism is out of the way, and the focus turns to his own dualist version, certain problems with the notion of zombies are brought to the fore via various thought experiments that were not hinted at earlier. I'm not saying there is any duplicity here—I have no doubt that Chalmers is sincere when he tells us that his arguments for the "natural" impossibility of zombies do not, in his view, undercut the notion that zombies are “logically possible”—yet the text bears some signs of stress. In the key thought-experiment of the dancing qualia Chalmers says that the bizarre possibility at issue seems “*only just* logically possible” (p. 269). One might think that logical possibility ought to be a clean-cut binary affair; the “only just” brings out the point that we are really talking about “what it is possible to imagine” (rather than any more precise logical concept), and it seems but a small step from “only just” conceivable to “not quite” conceivable. Something else: other philosophers have quite explicitly taken arguments very similar to those presented by Chalmers in chapter 7 as grounds for denying that zombies are conceivable. Chalmers registers this point, but—given the space he devotes to opposing arguments in other contexts—in an oddly perfunctory manner.[4]

In the same context—the rhetoric of the conceivable—one specific argument in chapter 3 is worthy of attention. At one point Chalmers, recognizing that his readers might have difficulty conceiving of a zombie, gives them a helpful nudge. His suggestion is that they might first try thinking about a *functional isomorph* of a normal human being, realized in a radically non-standard medium (as, for instance, Block’s “Chinese nation” idea). If you can successfully conceive of this weird isomorph as subjectively blank, he urges, then just transfer this blankness back to the physical duplicate of the ordinary human. Does this strike you as entirely kosher? To me, it partakes somewhat of the character of the following. “If you find it difficult to conceive of the idea that all of the money in the United States might be counterfeit, try this: first imagine that all the money in some third-world banana republic is counterfeit—pretty easy, isn’t it?—then reflect on the fact that there is no difference in principle between the monetary system of the US and the monetary system of the banana republic.” That is, the argument seems to come close to playing on the reader’s chauvinism. (In sharp contrast to Searle, for instance, Chalmers will go on to argue that any functional isomorph of a human being, no matter how outlandish the realization, will *in fact* have qualia.)

[4]A couple of sentences in the middle of p. 274, and notes 8 and 15 on pp. 387–8.

6 The etiology of belief in zombies

The observation above suggests a further point: In a case such as this, where it seems so difficult to find firm ground to stand on, we might make some headway in deciding whether or not we *ought* to believe in the possibility of X by attempting to unearth the *causes* of our inclination to believe (or disbelieve), insofar as these are different from the *reasons* we adduce and avow in argument. Chalmers makes a point along these lines, when he suggests that the profession among some philosophers of an inability to conceive of zombies actually flows from their *prior* desire to uphold a functionalist analysis of consciousness. This, he argues, has it backwards: the zombie question must be tackled on its merits, the chips falling where they may for materialist functionalism. Chalmers's observation has some force, but a case can also be made in the other direction, i.e., that a careful account of the etiology of belief in the conceivability of zombies may serve to undermine the attractiveness of the idea. I am not suggesting—in a parallel to Chalmers's point—that his profession of an ability to conceive of zombies is the product of a prior desire to reject materialism and sustain dualism (he tells us that this is far from the truth, and there is no reason to doubt him). The point is a different one, namely that the apparent (to some) conceivability of zombies may be an effect of an inappropriate choice of models or analogies.

The temptation to think that zombies are conceivable is, I suspect, based at least in part on the following thought. Human cognition and perception—as analysed in objective mode by the functionalists—is basically "just a lot of information-processing and behaviour modulation". But it is easy enough to conceive of information being processed and behaviour being modulated in the absence of any subjective, qualitative accompaniment. Think of a thermostat, the photocell at the grocery door (to repeat an earlier example), or your desktop PC. It is not hard to imagine these items doing their functional business without any subjectivity. The difficulty, if anything, lies in the opposite direction: the idea that my PC might enjoy a subjective phenomenology of mouse-clicks and key-presses, or of document-scans, does not come naturally.[5] A further source of temptation is the phenomenon of blindsight, a philosophers' favourite scientific finding if ever there was one. People with blindsight can discriminate, to some degree, states of the world presented to them in their blindfield (corresponding to a scotoma in visual cortex), and this is clearly accomplished via optical input, yet they say there is no subjective visual sensation associated with the blindfield. It is very tempting to inflate this possibility, to suppose that they *might* be able to discriminate as finely as ordinary people, but still without any visual phenomenology.[6]

If I have correctly diagnosed the sources of temptation, the "remedy" for a belief in zombies is just the sort of Dennettian exercise in imagination proposed above. One must be forced to recognize the huge gulf between the simple informational economies of the thermostat, and even the PC, and the amazingly subtle and layered informational economy of a normal human being. Taking the PC, or the *severely* degraded registrations of actual blindsight victims, as the model, one may fool oneself into thinking one has imagined something when one has not really confronted its detailed implications. This piece will have accomplished its aim if it encourages a few readers to take the latter possibility more seriously than hitherto.

References

Chalmers, David J., 1996, *The Conscious Mind: In Search of a Fundamental Theory*, New York and Oxford: Oxford University Press.

[5]Matters are a little complicated, in relation to Chalmers's argument, by the fact that he ends up (in chapter 8 of his book) giving serious consideration to a version of panpsychism, according to which such PC phenomenology might in fact exist. But Chalmers's attachment to this idea is quite tentative and circumspect, and I think he would agree that panpsychism holds no attraction *independent* of the dualist conclusions to which he finds himself driven by his prior conviction that zombies are logically possible, and that a satisfactory materialist account of qualia is impossible. In addition, I suspect that relatively few of his readers will be inclined wholeheartedly to follow in this particular direction.

[6]For what it is worth, this was my own temptation: at one time I did reckon I was able to conceive of zombies, with the facts of blindsight, as I understood them, providing the fuel. Also for what it is worth, my own conversion was by no means Pauline: I experienced a gradual shift of intuitions in favour of Dennett's position, to which I was at first very resistant. For an earlier stage in the journey, see Cottrell (1995).

Cottrell, Allin, 1995, "*Tertium Datur?* Reflections on Owen Flanagan's *Consciousness Reconsidered*," *Philosophical Psychology*, vol. 8, no. 1, pp. 85–103.

Damasio, Antonio, 1994, *Descartes' Error: Emotion, Reason, and the Human Brain*, New York: Grosset/Putnam.

Dennett, Daniel C., 1991, *Consciousness Explained*, Boston: Little, Brown.

Dennett, Daniel C., 1994, "Get Real," Reply to 14 essays, *Philosophical Topics*, vol. 22, no. 1 & 2, Spring & Fall, pp. 505–68.

Dennett, Daniel C., 1995a, "The Unimagined Preposterousness of Zombies," commentary on T. Moody, O. Flanagan and T. Polger, *Journal of Consciousness Studies*, vol. 2, no. 4, pp. 322–26.

Dennett, Daniel C., 1995b, Commentary on Ned Block, "On a Confusion about a Function of Consciousness," *Behavioral and Brain Sciences*, vol. 18, no. 2, pp. 252–53.

Dennett, Daniel C., 1996, *Kinds of Minds*, New York: Basic Books.

Wittgenstein, Ludwig, 1953, *Philosophical Investigations*, Oxford: Basil Blackwell.