

I, Zombie

[Penultimate Draft – December, 2000]

Paul Skokowski

Symbolic Systems

Stanford University

Stanford, CA, 94305

paulsko@stanford.edu

Abstract

It used to be that zombies were merely the stuff of horror fiction or Hollywood movies. But certain recent philosophical theories offer the prospect that zombies are possible. These theories argue that experiential contents, or qualia, are non-physical properties. The arguments are based on conceivability of alternate worlds in which physical laws and properties remain the same, but in which qualia either differ, or are absent altogether. This paper maintains that qualia are, on the contrary, physical properties in the world. It is shown how conceivability fails under the burden of the a posteriori identification of qualia with physical properties, and how a reasoned choice can be made between the two types of theories; which ultimately favors materialism, and rejects zombies.

In the *Monadology*, Leibniz asks us to “Suppose that there were a machine so constructed as to produce thought, feeling and perception, we could imagine it increased in size while retaining the same proportions, so that one could enter it as one might a mill. On going inside we should only see the parts impinging on one another; we should not see anything that would explain a perception.”(Leibniz, p. 181) The lesson to take away from Leibniz is this: You can’t find thought by peering into the brain. There is nothing red or leaf-shaped in the brain when you perceive a red leaf. You have to look elsewhere for the content. The same is true for experience. Mental events like thought or experience are mental because they are representational. I perceive objects in the world. I experience colors and shapes. Mental facts are mental because of their representational aspects.

Now, unlike Leibniz, I do not believe that the world is made up of simple substances, what he calls the monads, and their properties. I do not believe in pre-established harmony, or the re-creation of the simple substances by continual

fulgurations of the Divinity from moment to moment. No, I am a simple materialist. As such I believe the world is made of physical objects and their physical properties, together with nomic relations, or laws, that hold between certain types of objects under various conditions.

As a materialist, I would like to think that the phenomena of this world, including mental phenomena, can be explained in physical terms. Not likely, say opponents of materialism. One brass ring will forever remain beyond the grasp of physical explanation: conscious experience. There is a particular argument about the inadequacy of physical theory to explain conscious experience that I will examine in this paper. This is the dualist argument regarding the possibility of zombies, which finds a forceful proponent in David Chalmers.

A dualist uses arguments about zombies to establish that consciousness cannot be accounted for by materialism, or, to put it another way, that materialism about consciousness is false. Zombie arguments go something like this. Certain (perhaps all) mental states have a qualitative feel to them. Imagine a microphysical duplicate of me in another physical world that is itself a duplicate of this world. All the physical properties and all the physical laws which exist in our world also exist in the duplicate world. The zombie duplicate is not just like me molecule for molecule; he is also functionally just like me. Thus, given certain types of inputs he will produce certain types of outputs just as I would, based on the physical, and hence behavioral, duplication between the two worlds. In fact, this duplicate is like me in every way but one: he has no conscious experience at all. There is no subjective quality or character to his mental life. He has no conscious experience. He lacks qualia altogether. This duplicate is, in fact, a zombie.

Because my zombie duplicate is functionally identical to me, he has the same psychological or cognitive states. He has the same beliefs and desires, for example, that I do. He believes it is noon, and wants to go running. So he puts on his running shoes and outfit, and heads for the trails in the woods, where he doesn't have to dodge any cars, and the scenery is especially nice in the autumn. He gets tired towards the end, and winces when his ankle turns sharply on the edge of an unseen pothole. But the zombie has no phenomenal feel during his run. Though he responds in the appropriate ways to the colors¹ of the woods at this time of year, perhaps commenting to his running partner on the beauty of the big-leaf maples and black oaks against the backdrop of the sequoias, there is no internal color-*experience* to accompany the outward acknowledgement. There is no dull-feeling of pain to accompany his tiredness as he crests that last hill, or sharp blinding pain that accompanies the turning of his ankle, though both have their functional effects of slowing him down or making him limp for a few yards. Even the feeling of satisfaction, perhaps endorphin-induced, is lacking when he finishes, though he behaves as if he feels it.

The dualist tells us that it is perfectly coherent that my zombie duplicate can be like me in every way in this duplicate world, except that the zombie lacks conscious experience. He says that any story we tell about physical properties and processes in the zombie world will apply in our world, and vice-versa. If this is so, then physical processes will not account for consciousness, for that is what is lacking in the zombie-world. Facts about the consciousness that accompanies experiences in our world are further facts to those of physical theory.

¹ See Matthen 1988, Hilbert 1987, Tye 1995 on the color properties of objects. It doesn't matter whether the redness of a crayon turns out to be a monadic property, an ordered triple of spectral reflectances, or some other determinable set of properties, so long as this property is empirically measurable. Similarly for other objective properties, such as temperature, shape, and so on.

Thankfully, David Chalmers assures me that I live in *this* world, a world *with* consciousness, and, most likely, *without* zombies. How will I know this? Well, I *know* that I am conscious. It is my very experience of consciousness that provides the grounding of my knowledge. (Chalmers 1996, 103) This is no outside observation, rather, it is a first person experience of consciousness, an experience that Chalmers sometimes calls acquaintance. Having the experiences in question is what justifies our beliefs about consciousness. We are directly acquainted with these experiences. My feeling the pain on my run, or seeing the red leaves is what justifies my knowledge that I truly experience the phenomenal aspects of the pain or the color; that is, I truly experience pain, I am conscious of the pain; I truly experience red, I am conscious of red. This immediate phenomenal evidence is what rules out the possibility that I am a zombie. The zombie never has these experiences with these phenomenal properties. There is, in short, nothing it is like to be a zombie.

What am I acquainted with that my zombie is not? My qualia. For Chalmers qualia are the properties of the mental states we have which make them conscious states, or experiences. Thus, when I experience the red of the leaves of the big-leaf maple, the quale is a phenomenal, and a non-physical, property of my internal state, that internal state we call an experience. The experience, in so far as it is caused by, and causes, other physical states, may be thought of as a physical thing, while the qualia, the phenomenal quality of the state, is non-physical. It is an intrinsic, nonrelational property of the experience. Of course, if we subtract qualia from a world in which they exist, we still end up with beings who behave the same, who function the same. In this sense, non-physical qualia are epiphenomenal, and Chalmers admits that any theory of qualia remotely like his must face

up to this prospect.(Chalmers 1996, 158)

Now that we understand this style of zombie argument, I would like to return to examine materialism, particularly the sort of materialism that I subscribe to: externalism. As a materialist, I want to say that my experiential states are in my brain. But as an externalist, I want to point out that the *contents* of these states, as well as the contents of all my other quintessentially mental states, are generally to be found *outside* the brain. My mental states are, in short, intentional. They are about things, properties, and facts external to me. Back to experience, the contents of my experiences are properties in the world. Thus my experiences are about external properties. These properties are the contents of my experiences. These properties are qualia.

I stop on my autumn run and pick up an oak leaf which has caught my eye. The leaf is a spectacular bright red, with yellow-green highlights. Something is happening in my brain as I examine the leaf: sensory states which respond to color are active in my visual cortex. Something is also happening in the world as I examine the leaf: the leaf exemplifies the color properties of bright red, with yellow-green highlights. The colors I experience are properties of the leaf. The sensory visual state in my brain is the experience; the color is its content. (Nothing in my brain is a gorgeous bright red, with yellow-green highlights.) This experience, a color-experience, is a phenomenal state. Its content is color, and this color is found in the world, external to me. So the qualia of this experience -- its content, its redness, its yellow-greenishness -- ain't in the head.

Michael Tye is an ardent proponent of such an external materialism. For him, experiences are like other intentional states in that they have intentional contents. He designates the phenomenal contents of our experiences PANICs, where the acronym

stands for Poised Abstract Nonconceptual Intentional Content. The contents are poised in that they are the contents of the output representations of specialized “sensory modules, and stand ready and in position to make a direct impact on the belief desire system.”(Tye, 1995, p. 138) Thus these contents are poised to be used by our cognitive belief/desire systems. The contents are abstract in that what is represented at this level is not any particular concrete object, but rather some general property. The phenomenal content of a color experience is the color. The particular object which has that color is not part of the *phenomenal* content of that experience. My qualia is a red qualia, not a *the thing that is red* qualia. In fact the object may be absent. This is a form of misrepresentation. An example of how phenomenal contents are nonconceptual is afforded by considering colors. I may sense the color red₃₇ without having a concept of such a shade of red. I have a general concept of red, but not fine-grained concepts for determinate shades of red. My beliefs are limited by my concepts in a way my phenomenal experiences are not: my experiences of red₃₇ may perfectly well differ phenomenally from my experiences of red₂₅.

Fred Dretske is also a materialist about experience who is of the externalist persuasion. For Dretske, the senses have biological functions of providing information, and these indicator functions give rise to natural representations in an organism. There are two flavors of indicator functions and natural representations: *Systemic* indicator functions are inherited, or phylogenetic, and give rise to *systemic* representations. *Acquired* indicator functions are ontogenetic, acquired through learning, and give rise to *acquired* representations. Like Tye, Dretske points out there are also two sorts of mental representations: experience and thought. Experience is nonconceptual representation, and thought (beliefs, judgements, knowledge, etc.) is conceptual representation. Experiences

are states whose representational properties are *systemic*. Thoughts are states whose representational properties are *acquired*.

With Tye, Dretske identifies qualia with those properties that experience represents objects as having. The color blue is the property my experience systemically represents something as being; it is the content or quale of my experience of blue. The Representational Thesis which Dretske defends is a materialist thesis, so these properties are knowable objectively and the subject of a particular experience does not have privileged access to them. Qualia are objectively determinable because they are properties that the senses have natural functions to provide information about. These functions are as objectively determinable as the biological functions of bodily organs.

We now have enough at our disposal to compare the views of Chalmers and externalism. Suppose I look at a red oak leaf. I undergo a visual sensation in which I consciously experience red. Under Chalmers' analysis, there is a color property in the world, a physical property which the leaf exemplifies--a particular shade of red-- which I will call P. Further, I have a physical state in my brain, which we can call the experiential state E, which exemplifies physical properties covarying with the sensed property P. There is another aspect to my conscious experience of red. I am acquainted with -- I directly experience -- the phenomenal character or property, the *quale*, red. Call this property Q. Then Chalmers picture looks something like this: $E^Q(P)$, where E denotes the physical state in my brain together with its physical properties; P denotes the physical properties in the external world which are sensed when I experience red; and Q is the quale of my experience. We can see that the external property P causes state type E in my brain, and so my experience is about P. This allows Chalmers to explain my physical behavior on purely

causal, physical terms: I see the red leaf, which causes me to stop, pick it up, and comment on it to my running partner. The physical types E and P cover this sort of explanation, physical explanation, very well. But my experience has in addition a certain qualitative aspect, Q, which is what it is like for me to experience red. Q is the phenomenal content of the experience.

Now let us examine the externalist version of my undergoing a visual sensation in which I consciously experience red. In this world, there are no non-physical properties. Again, there is a color property in the world, a physical property which the leaf exemplifies--a particular shade of red-- which I will call P. Also, I have a physical state in my brain, which we can call the experiential state E, which exemplifies physical properties covarying with the sensed property P. And that is all. The color I experience is a property of the leaf. My experience is intentional. The internal part of the experience is a state in my brain, and the color is the content of this intentional state. Then the materialist picture looks like: E(P). This is a phenomenal state E, whose quale, P, is its content.

Suppose I inhabit such a world--a world physically identical to the world Chalmers describes, but which lacks non-physical properties. What conclusion would a dualist reach about me? I think it is painfully obvious (though it cannot be *my* pain): that I am a Zombie. For in the world I find myself in, I lack property Q, which leads directly to zombiehood. The fact that I lack property Q is something I and Chalmers can both agree on. We also agree that my physical behavior is based on nomic relations between physical properties in the brain and the external physical world.

Although it appears to a dualist that I am a zombie, we have overlooked something. When considering zombie duplicates, we normally accept micro-physical and functional

duplication, while subtracting out properties like Q. But how do I know which world I am in? As both Tye and Chalmers agree, it appears--at least on first glance--that one couldn't tell the difference. I will have the same *cognitive* states, including beliefs that I am seeing a red leaf, feeling pain in my left ankle, hearing Barber's Agnus Dei, and so on. As Tye puts it "what reason could possibly be given for supposing that I am not actually in" a zombie world?(Tye, 1995, 61) Given my beliefs and other cognitive states, how would I ever know the difference?

There seems to be two possible responses. The first we have already seen, which is that we are directly acquainted with the immediate phenomenal evidence of our experiences. Of course, having immediate acquaintance with qualia would rule out that I am a zombie. But hold on. It is clear to ME -- and this is an unshakeable brute fact as I stop to pick up the leaf -- *that I experience red now*. I do not just have a *belief* to this effect: I also *experience* red. And this experience is direct experience of a quality: I have an experiential state E in my brain whose content is the very property of redness P exemplified by the leaf. This content is the very phenomenal character, or quale, that counts as experiencing this shade of red. It appears to *me* that I am acquainted with this quale, for it is the very quale that I am experiencing now, regardless of my beliefs or other cognitive states. Since this *is* my experience, what could ever count as showing that this property of redness is not, in fact, a phenomenal content, is not, in fact a quale?

The dualist reply is that those who experience a quale Q have true beliefs to the effect that they experience Q, whereas I, zombie, do not have a true belief to this effect. Now, I grant that I do not have a true belief to the effect that I experience Q. But, I do have a true belief to the effect that I experience the property P, that is, that I experience red. Red

is a color quality. It is the content of my experience. It is a quale. It may not be a non-physical quale, but it is a quale nonetheless, and despite my zombie label, I can have true beliefs about it.

We could, at this point, give up on my zombie-characterization. We could say that, since I have a red quale and so am experiencing red, then I cannot be a zombie. Of course, in this view the externalist picture would be vindicated, for qualia would be a part of the physical world. However, since I lack property Q, I believe a dualist will continue to insist that I am a zombie: worlds without properties like Q are worlds without experience. Worlds with conscious experience require an acquaintance with Q via psycho-physical laws, that is, laws connecting non-physical with physical properties. To be sure, every physical property and physical relation remains the same as in zombie-duplicate worlds, but in addition, I have Q when I experience red. Since Q does not exist in the world I inhabit, I cannot be experiencing red.

For Chalmers and other dualists it is important to require property Q for experience because they argue we can imagine worlds in which the physical circumstances remain identical, but qualia are inverted. Perhaps in *this* world John experiences the quale red when confronted by a red object, but in an inverted world he experiences blue when confronted by a red object. He still behaves in every way as if he sees red in the inverted world, but his internal experience is of blue.

Note that what is essential to my experience of red is the content of the experience: red. I can pick out an experience of red in any world by this content. An experience of blue, whether in this world or another, is *essentially* different from an experience of red. Returning now to the world I inhabit, one which lacks non-physical properties, recall that

the color red that I experience is a property of the object I am sensing. The internal part of the experience E is a state in my brain, and the color P is the content of this intentional state, its quale. Since there is nothing more to the analysis, we can see that the color red is identical to the property P. Given this identity, we see that the property P is essential to my experience of red in this, or indeed in *any* possible world. That is, when I pick out an experience of red in any world I can imagine, I necessarily pick out an experience of the property P, since experiencing red *is* experiencing property P. There is no room to pick out any other color, since P is what red *is*.²

This shows that imaginability arguments such as spectral inversion won't work against externalism. Once the identity has been established, one can't imagine a world in which red wasn't this combination of physical properties. This will be true whether red turns out to be a particular N-tuple of reflectances (Tye, 1995), or whether it turns out to be whatever property it is the function of our red detectors to detect (Hilbert, 1992). This is no different from our inability to imagine a world in which water isn't H₂O, or in which (material) temperature isn't mean kinetic energy of molecules. On first glance it appears we can imagine such worlds: worlds with watery stuff, but where this stuff is XYZ and not H₂O. But we have discovered *a posteriori* that water *is* H₂O. This identity is necessary, and holds in every possible world. Therefore the XYZ that we refer to as "water" in some other world isn't water; in fact it's XYZ. Hence when we imagine water as something other than H₂O, we aren't imagining water at all. We are imagining some other substance. Similarly for the quale red. Once we've discovered that our quale red - the content of our experience

² Chalmers is compelled to agree, since, in his words, "...if something feels like a conscious experience, even in some counterfactual world, it is a conscious experience. All it means to be a conscious experience, in any possible world, is to have a certain feel."(Chalmers 1996, p. 133)

of red - *is* the physical property P in the world, then we have established an identity. When we try to imagine a world in which red is not the property P and in which we still experience red, all we will in fact be doing is imagining a world in which we experience some other property; some color other than red. Since red is essential to the experience of red, then we are not picking out the same experience in the other world.³

It appears that we have reached an impasse. A dualist and I will agree that I lack property Q. According to his characterization, I am a zombie. The dualist insists that that I lack qualia. But I insist that I do have qualia. How do we solve such a problem? Well, we can go to a mediator. How about William of Ockham?

Ockham's razor tells us not to multiply entities unnecessarily. The result is that the simplest explanation which accounts for all the phenomena is the best. There is plenty of agreement between the approach a dualist like Chalmers prefers and the one the externalist offers. Indeed, both agree completely on the relevance of physical properties and laws in explaining my behavior. The physical properties of my brain, physical properties in the external world, and nomic physical relations which connect them provide a basis which is embraced by both theories. Chalmers adds to this basis that we need an extra property, Q, in order to have conscious experience of a real physical property P, such as redness. Further, accounting for instances of properties like Q requires supplementary psychophysical laws -- in addition to the existing nomic relations in the physical world -- that will connect the two realms, nonphysical and physical.(Chalmers 1996) The externalist replies that the

³ A technical note. Chalmers holds that there are two intensions (meanings), which he calls primary and secondary intensions, attached to red (or water), and that the secondary (*a posteriori*) intensions are not relevant to picking out consciousness in other worlds.(Chalmers, 1996) But this distinction is a red herring for any externalist, since perceptual contents simply *are* properties of objects. Indeed, under a naturalistic account such as Dretske's (Dretske 1995) or Papineau's (Papineau, 1987), meaning is based on biological function, so primary intensions are otiose. They may serve a purpose as an explanatory device within philosophy of language, but naturalism has no need of them, and requires only physical properties of objects as the contents of our perceptual states.

physical properties of the world and the existing nomic relations are already doing the work of conscious experience. We don't need extra, nonphysical properties to account for qualia, for qualia are properties in the world that we experience. Q isn't the quale; P is, and P is a property exemplified in the world (and as such is a property within the mutually accepted physical basis of the world). And, of course, if we do not need nonphysical properties to account for qualia, we don't need psychophysical laws to account for their occurrences, or even the existence of a nonphysical realm at all!

Note that it is incumbent on dualism to show why the property Q is not epiphenomenal, though it is admitted, some measure of epiphenomenality must accompany dualist theories (Chalmers p. 158). But this quandary just doesn't occur for externalist theories. In externalism, an elegant economy of explanation is achieved beyond the lessening of entities and laws: the phenomena are directly linked in causal role between agent and world. The bottom line is, then, that the externalist has accounted for conscious experience with the physical tools on offer--and no more. Ockham's razor eliminates the epiphenomenal and non-physical dangles and chooses the simplest and most explanatory theory: externalism.

Chalmers will reply that the externalist, to use his phrase, has "not taken consciousness seriously". But how much more seriously can one take consciousness than to say that qualia exist as real physical properties, and are the contents of our experiences? Moreover what more satisfying answer could we come up with to what we experience when we experience red than "red"? This also provides a satisfying explanation for the picture that Leibniz drew earlier: we don't have to have red leaf-shaped objects in our brain in order to experience red. The red in the world will do perfectly well.

If one were a dualist, it might be tempting to say that beings who lack non-physical properties like Q and the psychophysical laws between realms that govern these properties are zombies. But this would be too hasty. Microphysical and functional duplicates of us living in a duplicate world will have conscious experience just like the beings that inhabit this world: *us*. Of course, this means that zombies, as duplicates without any conscious experience whatsoever, are impossible. So please don't bother weeping for your zombie duplicate; after all, she'll be feeling the same pain as you.

REFERENCES

- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences* **18**, 227-287.
- Chalmers, D. (1996). *The conscious mind*. Oxford: Oxford University Press.
- Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.
- Hilbert, D. (1992). What is Color Vision? *Philosophical Studies* 68:351-370.
- Hilbert, D. (1987). *Color and Color Perception*. Stanford:Stanford University/CSLI Press.
- Leibniz, G. (1973). The monadology. In G.H.R. Parkinson (Ed.), *Leibniz: Philosophical Writings*, pp. 179-194. London: J.M. Dent and Sons.
- Matthen, M. (1988). Biological functions and perceptual content. *Journal of Philosophy* **85**, 5-27.
- Papineau, D. (1987). *Reality and representation*. Oxford: Basil Blackwell.
- Tye, M. (1995). *Ten problems of consciousness*. Cambridge, MA: MIT Press.