

Chapter 22

Wittgenstein versus Zombies: An Investigation of Our Mental Concepts



Edward Witherspoon

22.1 Introduction

We human beings perceive our environment; we reflect on what we perceive; we have desires; we form intentions and act on them. All this mental activity is accompanied by what philosophers call ‘consciousness’, or the subjective experience of thinking. This is *what it’s like* to sense, reflect, desire, intend, and act. There is a distinctive inner feeling that goes with craving chocolate ice cream, and another that goes with seeing a purple armchair, and another that goes with wondering whether the weather will clear this afternoon.

I can be sure that I have consciousness, and I naturally suppose that the people around me do too. But some philosophers have argued that it is possible for there to be a creature that looks, acts, and talks just like a human being, but that lacks consciousness. They have appropriated the term ‘zombie’ as the name for such a hypothetical being. There is something that it is like to be me—a way my experience *feels*—and there is something very different that it is like to be a bat, but there is nothing it is like to be a zombie. A zombie is all dark inside. To make this possibility vivid, imagine a twin who is a molecule-for-molecule duplicate of you. It is logically possible, according to these philosophers, that your twin could lack the consciousness that you possess. In similar fashion, we can imagine a ‘zombie world’, viz., a world that is like ours in every physical respect, where every particle of matter is just as it is in our world, but in which there is no consciousness. I will call the claim that zombies are possible the ‘zombie hypothesis’ and philosophers who advance this claim ‘zombie theorists’.

The notion of a zombie has roots in early modern philosophy, but it came to prominence in the nineteen-seventies, when the possibility of zombies was used to argue against *physicalism*, a popular thesis in philosophy of mind. (This use of the

E. Witherspoon (✉)

Department of Philosophy, Colgate University, 13 Oak Drive, Hamilton, NY 13346, USA
e-mail: ewitherspoon@colgate.edu

zombie hypothesis receives its most developed expression in *The Conscious Mind*, by David Chalmers, and I will largely rely on this book as my source for what zombie theorists think.) Physicalism is the thesis that the physical facts determine all the facts, including all the facts about what we think and feel. The physical facts are the facts that belong to a completed physics, whatever that turns out to be; for our purposes, we may regard physical facts as the spatio-temporal disposition of the fundamental particles of matter, together with the laws of physics. Where Descartes leaves us with a conception of the mind as metaphysically and causally independent of matter, physicalism holds that the mind and all its features are ultimately determined by matter.

The zombie hypothesis challenges physicalism. For suppose there could be a zombie world that is particle-for-particle identical with our world, but that contains no consciousness. The physical facts (including, we may assume, the laws of physics) would be the same in both worlds, but since our world has consciousness while the zombie world would not, the physical facts do not determine the facts of consciousness. In other words, physicalism is false.

If physicalism is indeed false, its demise leaves an explanatory lacuna in the philosophy of mind: how is consciousness to be explained? It cannot be explained as a function of what is happening in the brain: the brain is a physical organ, completely describable in terms of the disposition of its neurons, and the falsity of physicalism entails that no physical facts (not even the physical facts about my brain) determine the facts of consciousness. Chalmers and many other philosophers of mind who follow him in rejecting physicalism have therefore cast about for other means for explaining the existence and character of consciousness. This problem has proven difficult, and so in philosophy of mind it goes by the quaint sobriquet, 'the hard problem'.

Like Chalmers, I am no fan of physicalism. Like him, I am interested in reflecting on the nature of consciousness. But in this essay I am not going to focus on either refuting physicalism or explaining consciousness, for I have trouble with the initial step of Chalmers's approach. I find it hard to accept the claim that zombies are possible. Indeed, I believe that claim is incoherent. What Chalmers and many others think is a description of a logical possibility turns out to be a piece of nonsense. This is a negative result, but from it we will be able to draw a positive lesson for how we should understand our mental concepts. This lesson will not solve the hard problem of consciousness, but it will perhaps make it seem less of a problem.

My inspiration for thinking through the problems with the zombie hypothesis is Ludwig Wittgenstein's *Philosophical Investigations*. Wittgenstein's major themes include the connections between thought, feeling, and action, and the tendency of philosophy to yield distorted descriptions of those connections. Wittgenstein can help us recognize that the zombie hypothesis rests on such a misdescription.

In what follows, I will explain more precisely just what a zombie is supposed to be. I will argue that the zombie hypothesis entails a form of skepticism about other minds. That in itself is not a definitive argument against the zombie hypothesis; indeed, Chalmers contrives to take a dose of skepticism in stride. Drawing on Wittgenstein, I will argue that the skeptical consequences of the zombie hypothesis are more

extensive than zombie theorists acknowledge. Even so, one might think that the possibility of zombies shows that skepticism is something that we have to live with. So I turn to another line of thought from Wittgenstein, about how the meaning of sensation language is established, to argue that the zombie hypothesis collapses into incoherence.¹

22.2 The Character of the Possibility of Zombies

Let's return to the key claim, that zombies are possible. Chalmers presents it thus:

How, for example, would one argue that a mile-high unicycle is logically possible? It just seems obvious. Although no such thing exists in the real world, the description certainly appears to be coherent. ...

I confess that the logical possibility of zombies seems equally obvious to me. A zombie is just something physically identical to me, but which has no conscious experience – all is dark inside. While this is probably empirically impossible, it certainly seems that a coherent situation is described; I can discern no contradiction in the description. ... Almost everybody, it seems to me, is capable of conceiving of this possibility. (Chalmers 1996, p. 96)

When a zombie theorist says that zombies are possible, that does not mean that they are *empirically* possible or *physically* possible. The relevant notion of possibility instead is that of *logical* possibility. We can illustrate the differences between these notions of possibility with some examples. It is logically possible that there is an island of pure gold rising out of the seabed in the Pacific Ocean. It is physically possible as well, in that the existence of such an island would violate no laws of physics. But it is not empirically possible, in that there is not enough gold in the vicinity of earth to form such a mountain; such a mountain cannot exist in the actual world. It is logically possible that there is a unicycle one mile tall, but such a unicycle is physically impossible, because no possible material could be rigid enough yet also slender and light enough to construct one; and since it is physically impossible it is *eo ipso* empirically impossible. But a golden mountain and a mile-high unicycle are *logically* possible because the descriptions of these things entail no contradictions. So when philosophers say that a zombie is possible, they are saying that there is no contradiction in the idea of a being that is identical to a conscious human being in every physical respect but that nonetheless lacks consciousness; in other words, we can give a coherent description of such a being.

Chalmers finds that the logical possibility of zombies is obvious. I confess that I find the logical possibility of zombies far from obvious. Perhaps this is because I have been influenced by passages like this one from Wittgenstein's *Investigations*:

But can't I imagine that people around me are automata, lack consciousness, even though they behave in the same way as usual? — If I imagine it now – alone in my room – I see people with fixed looks (as in a trance) going about their business – the idea is perhaps a

¹Eric Marcus reaches a similar conclusion via a quite different argument in "Why Zombies are Inconceivable" (Marcus 2004).

little uncanny. But just try to hang onto 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 empty; or you will produce in yourself some kind of uncanny feeling, or something of the sort. ... (PI, §420)

Wittgenstein tries to imagine that the people around him lack consciousness (*Bewusstsein*). He asks us to try saying, about some frolicking children, ‘Their minds are all dark inside; they feel nothing.’ When I try this I find, as Wittgenstein predicts, that those words start to feel empty; if I nevertheless hang onto those words, I find my attitudes towards the children become unstable. In either case, the children’s being zombies is *not* a possibility I can easily imagine or conceive.

Even if others share this experience, we are not yet at a useful result, for our experience may be a merely psychological feature of me and like-minded thinkers. If zombie theorists find it easy to imagine zombies, perhaps they just have different psychological powers—greater powers of imagination, perhaps. Wittgenstein’s remark does not settle the question of whether zombies are logically possible. We need to find some additional argumentative resources. We will start this search by considering the mental states zombies *do* have, according to zombie theorists.

22.3 Psychological and Phenomenal Mental Concepts

A zombie theorist holds that a being physically identical to me—or physically identical to a frolicking child—could lack consciousness. But still there has to be some sense in which even a zombie is conscious: for zombies are sometimes awake (and conscious of their surroundings) and sometimes asleep (and unconscious). Moreover, zombies have to be able to (in some sense) perceive and learn and plan, for otherwise they would behave very differently from human beings and so would fail to be physically identical to them.

Accordingly, zombie theorists allow that a zombie does have mental states. A zombie can perceive, can learn, can form intentions and act on them. In short, it can think and have experiences. When zombie theorists say that a zombie lacks consciousness, they do not mean that a zombie is unconscious in the way someone drugged or asleep is unconscious. An awake zombie possesses consciousness in that it is aware of and responds to its environment. But its awareness and responsiveness lack the subjective ‘experience’ of thinking; it doesn’t *feel* anything. For us human beings, there is a characteristic feeling of joy, for example. When *we* see a loved one after an absence, there is a special thrill of reunion that accompanies our smiles and hugs. A zombie lacks all these feelings (despite recognizing a ‘loved one’ and smiling and embracing them just as you or I would).

So a zombie has experience and consciousness, in that it is aware of and thinks about its environment; but it lacks experience and consciousness in that it lacks the subjective, inner *feelings* that belong to our awareness and our thought. To forestall possible confusion arising from these ways of describing a zombie, Chalmers dis-

tinguishes two categories of mental concepts: psychological and phenomenal. The psychological concepts apply to the mind viewed as a causal nexus. These concepts are defined in terms of what produces them and what effects they have. The mind conceived in terms of psychological concepts is an information-processing organ that mediates between inputs (e.g., sensations) and outputs (actions); it infers, reflects, and learns. (A zombie's learning is the modification of its information-processing pathways so that its future behavior enables it to cope better with its environment.) Psychological concepts include concepts such as perception, inference, intention, belief, and desire. They apply to an agent's mind from a third-person perspective; that is, the attribution of these concepts is based on observation of a creature's interaction with its environment.

Phenomenal concepts belong to the mind considered as the locus of feeling. Consciousness in the phenomenal sense is *what it is like* to have experience. (I will follow the usage of zombie theorists in using the term 'consciousness'—without modification—to refer to consciousness in the phenomenal sense.) Phenomenal concepts apply to an agent from a first-person perspective, in that they concern what experience is like *for me*. Phenomenal concepts include the *qualia* of pain, the feeling of joy, the sensation of red, and the *feels* that accompany thinking, willing, desiring, and so on.

Every mental property, Chalmers argues, is either psychological, phenomenal, or partakes of both categories. For example, color perception is a psychological concept; the *qualia* or *look* of the color is a phenomenal concept. Color perception is a key part of the explanation of how both human beings and zombies respond differentially to (for example) red and green traffic lights. But zombies have no *qualia*: their experience does not extend to the *look* of the color or the *feel* of a piece of velvet or the *taste* of a pineapple. *Pain* is an example of a concept that has both psychological and phenomenal dimensions. As Chalmers puts it:

The term [sc. 'pain'] is often used to name a particular sort of unpleasant phenomenal quality, in which case a phenomenal notion is central. But there is also a psychological notion associated with the term: roughly, the sort of state that tends to be produced by damage to the organism, tends to lead to aversion reactions, and so on. (Chalmers 1996, 17)

We will return to this idea that the word 'pain' really encompasses two distinct notions: a psychological information state whose content is that the organism is probably being damaged, and a phenomenal state whose content is the painful feeling itself. The distinction is more problematic than zombie theorists realize.

22.4 Zombies and the Problem of Other Minds

I mentioned above that the zombie hypothesis has recently been used to argue against physicalism in the philosophy of mind. But the description of zombies can also be seen as an updated expression of a much older problem, that of skepticism about other

minds. (Compare the use of brain-in-a-vat scenarios to provide a vivid, contemporary portrayal of skepticism about the external world.) Skepticism about other minds has various versions, but the characteristically modern form of it holds that all the information I could ever get about other people (their words and behavior, their history, even the workings of their brain and nervous system) does not enable me to know what they are thinking or what their experiences are. Skepticism represents the other's body as a curtain that drops between my mind and hers: the body emits what sound like groans of pain or exhibits what appears to be a smile of joy, but the groans could be pretense or the smile could be a mere reflex act that expresses nothing.²

There are obvious affinities between the zombie hypothesis and skepticism about other minds. But it is worth making some of the connections explicit, because this will show some of the philosophical costs of accepting the zombie hypothesis. I will argue that the zombie hypothesis entails skepticism about other minds. It will turn out that, for a zombie theorist, there is no justification for believing that anyone besides oneself has consciousness in the phenomenal sense. This may come as a surprise, given the fact that the zombie hypothesis is one premise in an argument against physicalism whose other main premise is that there are facts of phenomenal consciousness, that is, that human beings (the physical twins of zombies) are conscious. But this anti-physicalist argument only requires the premise that *I* (the proponent of the argument) am conscious in the phenomenal sense. It would work even if I doubted the existence of any other consciousnesses.

The connection between the zombie hypothesis and the doubtfulness of consciousnesses other than my own can be seen in Chalmers's own discussion of skepticism about other minds. He entertains a thesis called 'eliminativism', which is the view that, really, no creatures are conscious, despite our tendency to attribute feelings to ourselves and others. He writes:

Eliminativism about conscious experience is an unreasonable position *only* because of our own acquaintance with it. ... To put it another way, there is an *epistemic asymmetry* in our knowledge of consciousness that is not present in our knowledge of other phenomena. Our knowledge that conscious experience exists derives primarily from our own case, with external evidence playing at best a secondary role.

The point can also be made by pointing to the existence of a problem of other minds. Even when we know everything physical about other creatures, we do not *know* for certain that they are conscious, or what their experiences are (although we may have good reason to believe that they are). (Chalmers 1996, 102)

Chalmers is saying that I know only that *I* have consciousness. My phenomenal experience is the only basis for my belief that consciousness exists. When it comes to other creatures, I do not know what their phenomenal experiences are; I do not even know whether they have experiences. At best, I have 'good reason' to believe that they do.

²One could press the question of antecedents, to ask what gave rise to skepticism about other minds and why that problem became acute in the nineteenth century and remains so today. (For a brief synopsis of the history of the problem, see Hyslop (2018). These important questions lie beyond the scope of the present essay.

This is not a merely incidental remark of Chalmers. Anyone who accepts the zombie hypothesis is committed to skepticism about other minds. To see this, reflect on the thesis that there could be two physically identical creatures, one of whom possessed consciousness while the other one lacked it. So let us assume that this is the case, and suppose further that I encounter these two identical humanoids.³ (We may add, if you like, that the creatures have some properties that enable me tell them apart, but that are clearly irrelevant to whether they possess consciousness: let's suppose that they are in different rooms within my thought-experimental laboratory and are wearing different-colored wristbands.) Could I ever know which one of the two humanoids before me has consciousness, and which one doesn't? This is clearly a case in which whatever I know about the others will be based on my perception of them. If the range of my perceptions is limited to physical properties, then it follows straightforwardly that I cannot perceive any difference between the person (with phenomenal experience) and the zombie (without it), since *ex hypothesi* they have the same physical properties, except for whatever trivial properties allow me to distinguish one individual from the other. It follows that I cannot know whether any humanoid that I encounter is a zombie or a human being; and if I cannot know that, then I certainly cannot know anything about the character of the consciousness of the being before me.

This argument depends on the premise that my perception of others is limited to their physical properties. Could it happen instead that I perceive another's phenomenal consciousness directly, without my perception being mediated by a physical state of the other? Chalmers's distinction between psychological and phenomenal elements of the mind entails that the answer is no. My perceptions are *psychological* features of me and are therefore entirely dependent on the physical properties of my sense organs and the stimuli they receive. (I and my zombie twin have the same perceptions.) If we were to suppose that I could perceive another human's consciousness directly (without a physical intermediary), then the other's consciousness would be making a difference to my physical properties without itself being physical, and that would violate Chalmers's position that the physical realm is causally closed.⁴

Hence we may conclude that, if zombies are possible, I cannot know anything about the phenomenal character of another's mental life. We can make this conclusion vivid through the well-known idea of an 'inverted spectrum'.⁵ For all I can tell, another person may have color experience (in the sense of the phenomenal *qualia* or *feel* that colored objects evoke in her) that is systematically different from mine. The objects that give her the subjective qualia of *blueness* give me the qualia of *redness*,

³I use the term 'humanoid' to refer to a creature that is either a human being (who possesses consciousness) or a zombie (who doesn't).

⁴The best evidence of contemporary science tells us that the physical world is more or less causally closed: for every physical event, there is a physical sufficient cause. If so, there is no room for a mental "ghost in the machine" to do any extra causal work.' The qualification 'more or less' is to allow for quantum indeterminacy which, Chalmers argues, 'cannot be exploited to yield a causal role for a nonphysical mind' (Chalmers 1996, p. 125).

⁵Chalmers embraces the possibility of an inverted spectrum and uses it as an additional argument against physicalism at (Chalmers 1996, pp. 99–101).

and things that look green to her look yellow to me, and so forth throughout the color spectrum. If, for each of us, our qualia are consistently aligned with colored objects, we will use the same color words for the same objects. We will both say that the setting sun is ‘red’, even though the person whose subjective color spectrum is inverted with respect to mine is having the experience that, if it were mine, I would call ‘blue’, while I of course am having the experience that I call ‘red’. There is no way for me to know what another person’s color experience is like, for these possible differences can never show up for us. The skeptical conclusion we have reached from the zombie hypothesis generalizes this result for all our phenomenal experiences. The zombie hypothesis implies that I can never know what experiences another person has, or even that the humanoids around me *have* experiences. For all I know, my friends and family might all be zombies—thinking, talking, reasoning zombies, but beings that lack consciousness nonetheless.

One might think that this would be a disturbing consequence that should cause philosophers to rethink their commitment to the zombie hypothesis. But zombie theorists have a way to make peace with skepticism. Their way of accommodating themselves to skepticism about other minds is hinted at in the Chalmers passage quoted above: although I cannot *know* whether others have consciousness, I can have *good reason to believe* that they do, and that their experiences align with mine. In other words, I am justified in believing that other humanoids are indeed human beings. Perhaps zombie theorists can be content to give up knowledge claims about others’ experiences because they can rely on well-grounded beliefs about them.

Indeed, it is not unreasonable to hold that well-grounded beliefs—not knowledge—are sufficient for our practical needs and for sustaining our relationships with others. One could argue that *knowledge* makes demands for certainty that we don’t have to satisfy in order to live a rationally ordered and meaningful life.⁶ Do I have to *know* that my car is in the parking lot in order for me to act rationally in walking to the lot in order to get home? Isn’t it enough for me to have a well-grounded belief that it is there? (Do I even know that my car is in the lot? Do I know that it has not been stolen? Or towed away? Or swallowed by a sinkhole?) Similarly, I could base my relationships with others on a well-grounded belief that our phenomenal experiences are similar. Such beliefs could be reassure me, for example, that my loved ones have feelings too.

But *do* I have good reasons for believing that others have phenomenal experiences similar to mine? For a zombie theorist, any attempt to provide such reasons will have to start with my own conscious experience. (As Chalmers writes, ‘Our knowledge that conscious experience exists derives primarily from our own case’ (Chalmers 1996, p. 102).) From the facts about my consciousness, together with any information I can obtain about the physical attributes of myself and others, I will need to construct a reasonable basis for ascribing a similar consciousness to others. The resulting

⁶I do not mean to endorse this line of argument. A contrary position is that the concept of knowledge is fundamental, and that *merely having a justified belief that p* should be understood as a privation of the fundamental relation of *knowing that p*. Investigating this issue, while important, would take us beyond the scope of this essay.

argument is usually called ‘the argument from analogy’. It runs as follows: *I* am conscious, and my consciousness is associated with certain physical facts (e.g., facts about environmental impingements on my sensory organs, facts about the sounds coming out of my mouth, facts about my movements). When I observe other beings receiving relevantly similar impingements, making similar sounds, moving in similar ways, then I may infer that they are conscious too. The behavior of other humanoids is similar to mine; my behavior is accompanied by consciousness; therefore, their behavior is accompanied by similar consciousness as well. This argument appears to give me good reasons for believing what I am naturally inclined to believe anyway, viz., that others have phenomenal experiences like mine.

But now I want to bring Wittgenstein to bear to argue that the zombie hypothesis has the consequence, not only that I don’t *know* about another creature’s consciousness, but also that I don’t have any *well-grounded beliefs* about it either. This is a much more troubling consequence than Chalmers’s official position that I don’t have knowledge but do have justified beliefs about others’ consciousnesses.

22.5 Wittgenstein on Skepticism About Other Minds

Wittgenstein thinks that the argument from analogy is hopeless. He thinks it has two flaws. First, the inductive base (the data about my own consciousness and its connection to physical properties), from which I am supposed to infer features of the consciousness of others, does not exist. Second, even if I did have such data about myself, extrapolating that data to a conclusion about others’ consciousnesses would be an inductively weak generalization.

Let’s consider the data from my own case from which I am supposed to generalize to others’. This is supposed to be the set of correlations I have observed between my states of consciousness and physical facts about me and my environment. For example, I feel a burning sensation when I grab a pot on the stove, and I immediately yell and pull my hand away. According to the argument, this provides me with a data pair: the feeling of burning is associated with a certain type of stimulus and physical movement and a characteristic noise (a yelled ‘ouch’). When I get enough of these data pairs in my data set, then when I observe someone else jerking her hand away from a hot pot and yelling ‘ouch!’, I can justifiably infer that she feels a burning pain.

The first problem with this argument is that my relation to my own behavior is not that of observer to observed. My awareness of what I am doing normally comes from my doing itself, not from my observation of my doing. And when I act, I already think of my behavior as caused by or expressive of my consciousness. I jerk my hand away because of the pain; my smile expresses my contentment, as my weeping expresses my sadness or dismay. It is a difficult mental exercise (one which most people rarely perform) to separate out the physical features of one’s action from its intimate connection with consciousness and to regard these as independently describable elements; but such a separation is just what one has to do to in order to construct the required database. And while I can sometimes observe myself acting,

doing so requires special circumstances and a stance towards myself quite different from my normal engaged perspective. The very specialness of these circumstances shows that we do not have the resources to construct a database of correlations between my observed behaviors and my phenomenal experience.

An especially crippling gap in this putative data set concerns facial expressions. For proponents of the argument from analogy, reading others' facial expressions would have to play a huge role in my attributions of feelings to them. But I hardly ever observe my own facial expressions. (In the days before mirrors and mobile phone selfies, we could safely say that people almost never observed their own facial expressions.) So we cannot construct the data set that, according to this argument, is required for the vast majority of attributions of feelings.

But now even if I could construct a database of the correlations between my consciousness and physical facts about my behavior, the argument from analogy would run into a second problem. For my database would contain those correlations for just one humanoid: me. (I cannot include observations regarding the correlations between consciousness and its physical manifestations from any other humanoids, for that would assume the conclusion that the argument is trying to justify.) It would be an extreme case of hasty generalization for me to use the fact of a correlation between *my* behavior and *my* consciousness to infer a correlation between other humanoids' behavior and their (alleged) consciousnesses. This idea is captured in Wittgenstein's remark in PI §293: "If I say of myself that it is only from my own case that I know what the word "pain" means—must I not say *that* of other people too? And how can I generalize the *one* case so irresponsibly?"

Finally, in addition to criticizing the argument from analogy, Wittgenstein notes that we do not actually employ any inference such as it posits:

"You say you attend to a man who groans because experience has taught you that you yourself groan when you feel such-and-such. But as you don't in fact make any such inference, we can abandon the justification by analogy." (*Zettel* §537)

That we treat other humanoids as conscious human beings is a fundamental fact about us: our so treating them is not justified by argument. That is not a rational defect on our part, because our attitude towards them does not stand in need of justification.

Wittgenstein has shown that the cost of accepting the possibility of zombies is higher than Chalmers thinks it is, for that possibility entails skepticism about other minds, in a radical form according to which we can have no well-grounded beliefs that others are conscious. But someone who thinks it is *obvious* that zombies are possible could take this in stride. Skepticism about other minds might just be our human plight. In other words, Wittgensteinian arguments show that the zombie hypothesis comes at a very high cost. But if you think that zombies are possible, you might decide that you simply have to pay that cost: you must give up any claim to rational beliefs that there are conscious beings around you. You might have to tolerate the existence of widespread prejudices that others are conscious, for ingrained habits of thought tend to linger, even after we have recognized that they are groundless. But the idea that others are conscious will be mere prejudice, and rationally scrupulous

thinkers will carry on their lives as much as possible without ascribing consciousness to other beings.

Wittgenstein offers some striking statements of how difficult it is to sustain such a stance. “Just try – in a real case – to doubt someone else’s fear or pain!” (PI, §303). But one might think that these remarks merely express the *psychological* difficulty of giving up ingrained habits of thought. They don’t suggest—nor does the preceding critique of the argument from analogy entail—that zombies are *not* logically possible.

To deepen my Wittgensteinian criticism of the idea of a zombie, I will argue that the zombie hypothesis is incoherent, that it only *appears* to be a philosophical thesis. This is a very different argumentative tack from the preceding one, in which I have been treating the zombie hypothesis as a thesis with logical consequences (e.g., that one can never have a justified belief about the consciousness of another). In what follows, I will use Wittgenstein’s ideas to argue that zombie theorists deprive themselves of the resources required to make sense of the language of phenomenal concepts.

22.6 The Meaninglessness of the Attempt to State the Zombie Hypothesis

The problem I find in the zombie hypothesis is that it makes it impossible to account for the acquisition of a language for phenomenal concepts. To see why, we need to recall the distinction Chalmers makes between the two kinds of mental concepts, the psychological concepts that apply to both zombies and me, and the phenomenal ones that apply to me but not to zombies. Let us recall that a psychological mental concept applies to the mind considered as an information-processing organ; such concepts are grounded in the network of causal relations between environmental stimulus and the behavior of the organism. A phenomenal mental concept applies to the mind as a locus of feeling. According to Chalmers, many of our ordinary mental concepts partake of both aspects: even the all-important concept of consciousness can be subdivided into a psychological component and a phenomenal component.

Our ordinary concept of *pain* is another case of an amalgam of two distinct components, according to zombie theorists. Pain in the psychological sense is a causal intermediary between injury to the organism and aversive behavior; it is the concept that applies when we explain why an organism pulls away from a hot fire. Pain in the phenomenal sense is the distinctive feeling that (in human beings but not in zombies) accompanies injurious stimuli and aversive responses.

As a shorthand, instead of talking about pain in the phenomenal sense, I will refer to ‘felt pain’. Now this expression, ‘felt pain’ and other phenomenal ones, like ‘hurts’, ‘feels good’, ‘tastes like vanilla’ are required in order to state the zombie hypothesis: a zombie is a creature that is physically just like me, but that does not have the felt pains, the felt pleasures, and the rest of the qualia that I have. Since the zombie hypothesis entails that anything I know about phenomenal experience is

derived from my own case, when I learn expressions like ‘felt pain’ I can draw only on my own resources.

How do I learn how to use the expression ‘felt pain’? The story will have to go something like this. Something happens to me: I stub my toe, say. This event is accompanied by an experience in the phenomenal sense. I decide to label this experience ‘felt pain’. Later, something else happens: I twist my ankle. This event is accompanied by another experience, which I judge to be relevantly similar to the experience I remember from when I stubbed my toe. So I say that I am having another felt pain, this time in my ankle instead of my toe.

This account of learning phenomenal language sounds simple and straightforward. But Wittgenstein shows us that it is anything but. Let’s focus on a key step of the above account: I have a sensation, and I focus my attention on it and dub it ‘felt pain’. What is involved in this? This act of establishing the meaning of my expression ‘felt pain’ has to set up a rule. The rule will determine whether my future applications of those words is correct or incorrect. (It is a necessary feature of words’ having a meaning that uses of them are normatively constrained, that is, that a speaker’s use of them can be correct or incorrect.)

Wittgenstein asks, how do I establish this rule for myself? One might think that there is nothing to it: I mentally focus on the pain, and set up the rule that my expression ‘felt pain’ will refer to *that*. Now for this rule to govern my future use of the words ‘felt pain’, I have to have some criterion for determining that a future sensation is the same as the one I am having now. At first blush, this presents no special problem: the criterion can be my memory of the experience. When I twist my ankle some days later, I recall my earlier experience of felt pain and compare my present sensation to it so as to judge that my present sensation is relevantly similar to my earlier one.

Wittgenstein would draw our attention to two conditions that the zombie hypothesis imposes on this account. First, the criterion I employ is one which I alone possess and which I must be able to apply without any check from others. For if the humanoids around me are zombies, whatever words they come out with will not get their meaning from a connection with a felt pain (as mine does), so anything they say about the correctness (or not) of my use of my words ‘felt pain’ will not have any bearing on what *I* am using as the criterion of correctness. And I cannot assume that any of the humanoids around me are people, *not* zombies, because, as we have seen, it would be unreasonable for me to believe that. Moreover, what we are trying to explain is how I am able to meaningfully say, ‘So-and-so is not a zombie but has felt pains and other phenomenal states’; so at this point in the argument, I cannot help myself to the conceptual resources required to formulate the assumption.

A second feature of the above account of how I learn the expression ‘felt pain’ is that it relies on my *memory* of the earlier pain, since that memory becomes the criterion that determines the correctness of later applications of the expression ‘felt pain’. Wittgenstein points out that if I am remembering my earlier pain (from when I banged my toe) when I later say again ‘I feel pain’, then there is a possibility that I am *misremembering* it. Suppose I do misremember, and now use ‘felt pain’ for a sensation that in the past I would have labelled ‘an itch’. If this happened, if my

memory deceived me, I would have no means to detect the deceit. I cannot rely on anything that other humanoids say. And I cannot rely on the causal nexus between my own sensory inputs and typical behavior to reason that this is the kind of injury that typically results in *pain* and not itches; such reasoning could establish only that the psychological concept of pain applies to me, but the zombie theorist has severed the tie between the psychological concept and the phenomenal concept, so I cannot infer from the fact of my injury that I am feeling pain, not an itch.

So when I use the expression ‘felt pain’, I have nothing to rely on except my memory, and nothing against which I can check my memory. In other words, I am relying only on what I *seem* to remember. If my present experience *seems to me now* like a pain, then I will call it a pain. But in that case, I don’t actually have a criterion for the correct use of the words ‘felt pain’; for all I can tell, I might be applying those words to an itch that my memory has mistakenly brought up as the criterion of application for the words ‘felt pain’.

The argument that I have been developing is expressed by Wittgenstein in a passage about an attempt to define a sign ‘S’ as the name of a certain sensation. Wittgenstein describes how one might try to ostensively define ‘S’. (An ostensive definition is one in which I pick out what a word means by pointing to the object it describes or refers to.)

Can I point to the sensation? – Not in the ordinary sense. But I speak, or write the sign down, and at the same time I concentrate my attention on the sensation – and so, as it were, point to it inwardly. – But what is this ceremony for? For that is all it seems to be! A definition serves to lay down the meaning of a sign, doesn’t it? — Well, that is done precisely by concentrating my attention; for in this way I commit to memory the connection between the sign and the sensation. – But “I commit it to memory” can only mean: the process brings it about that I remember the connection *correctly* in the future. But in the present case, I have no criterion of correctness. One would like to say: whatever is going to seem correct to me is correct. And that only means that here we can’t talk about ‘correct’. (PI, §258)

The account the zombie theorist has to give of how we learn expressions for phenomenal concepts leaves us with no criterion for the correct application of those expressions: whatever is going to seem correct to me at the moment I use them is correct, which is as much as to say that the notion of correctness does not apply. And if I have no criterion of correctness for these words, they don’t mean anything. The language of phenomenal experience (‘felt pain’, ‘sensations’, and all the rest) is, in the mouth of a zombie theorist, nonsense.

Here is another way to present Wittgenstein’s argument. In order to establish a meaning for the expression ‘felt pain’, I resolve to use those words for the feeling I am having now, when I stub my toe. And I resolve to use those words *only* for this feeling, whenever it may recur. I am *not* resolving to use the words for whatever will at some future time *seem* to me then to be this feeling. For suppose that, in the moment when I resolve to use the words ‘felt pain’ for the feeling I am currently having, I also imagine that my memory were to go haywire in such a way that I would in the future think that the words ‘felt pain’ named the feeling that I now call ‘an itch’. I would not regard it as acceptable to use the words ‘felt pain’ in the future for that itch; I would not regard that haywire use of words as correct or as in line with

my intention for using ‘felt pain’. My memory on its own cannot distinguish what I will then mistakenly regard as *felt pain* from what I now label ‘felt pain’. So my memory cannot provide a criterion of correct application of the words, and so cannot yield a rule for the use of the expression ‘felt pain’. But a consequence of the zombie hypothesis is that the only resource I have for guiding my future use of the words is my memory. And this implies that (given the commitments of a zombie theorist) I lack the resources for framing a rule for my use of the expression ‘felt pain’, which is to say that I cannot assign a meaning to those words.

In giving this argument, I have taken as an example of phenomenal language the expression ‘felt pain’, which I have stipulated to be synonymous with the phenomenal dimension of the concept of pain. But there is nothing special about *felt pain* as compared to any other phenomenal concept. So this argument establishes that any expression that zombie theorists purport to use to refer to their phenomenal consciousnesses is meaningless.

This prepares us to locate an incoherence in zombie theorists’ attempt to formulate their hypothesis. This emerges most clearly when we take a first-person perspective on the attempt. I start, then, by saying that it is logically possible for there to be creatures that are like me in every physical respect but that lack phenomenal consciousness. A consequence of this is that I have no good reason for thinking that anyone else is a human being (with consciousness) rather than a zombie (without). This may be disturbing. But at least I know that *I* have consciousness, don’t I? Don’t I know that I, unlike zombies, have felt pains, sensations, feelings of all sorts? But what do I mean by the expression ‘felt pain’? *I* cannot give it a meaning (according to the argument we have just seen), and I cannot draw on a meaning provided from elsewhere. The same goes for any other expression that purports to apply to phenomenal concepts, including the concept *phenomenal consciousness* itself: in my mouth, all those expressions are nonsense. And so when I (the would-be zombie theorist) come out with the words ‘I have consciousness, but a zombie does not’, my words have no meaning. What looked like a sentence expressing a hypothesis to the effect that zombies are possible turns out to be nonsense.⁷

22.7 Conclusion

This argument, if it is successful, exemplifies the method that Wittgenstein describes in PI §464: “What I want to teach is: to pass from unobvious nonsense to obvious nonsense.” As I see it, the zombie theorist traffics in unobvious nonsense. What has emerged is that there are logical conditions for the meaningful employment of phenomenal language that the zombie theorist cannot meet. The zombie theorist is com-

⁷Robert Kirk asserts that using Wittgenstein’s private language argument against the conceivability of zombies will amount to begging the question against those who think zombies are conceivable (Kirk 2019). I hope that the structure of my presentation shows that a Wittgensteinian argument does not have to beg the question.

mitted to establishing the meaning of phenomenal language using only the resources of his own feelings and memories. But these resources are not sufficient to enable the theorist to give a meaning to phenomenal language. And without that phenomenal language, the attempt to state the zombie hypothesis results in a nonsense utterance. The zombie theorist only appears to be stating a thesis. Consequently the zombie theorist does not present us with a possibility that refutes physicalism; the zombie theorist does not saddle us with the ‘hard problem’ of explaining consciousness.

This criticism of zombie theorists appears to be in tension with my earlier argument that the zombie hypothesis entails a radical form of skepticism about other minds. This cannot be an appropriate description of the earlier argument, since we have just concluded that the attempt to state the zombie hypothesis results in nonsense, and a piece of nonsense cannot have any logical implications. How then should we characterize the status of that earlier argument? Similarly, how should we characterize the intermediate steps of the argument of the preceding section, in which I have sometimes used expressions like ‘a consequence of the zombie hypothesis’? These are not easy questions, but I think we can begin to address them by drawing an analogy between the zombie theorist and someone who has inconsistent beliefs without being aware of their inconsistency. So long as the subject is thinking about each belief in isolation from those it is inconsistent with, she can draw conclusions from it. She is stymied in her cognition only when she brings the inconsistent beliefs together in thought in a way that manifests their incoherence: at that point, she will not be able to retain all of them, and some of the inferences that her former beliefs supported will become unjustified. In a similar fashion, so long as we (and the zombie theorist) are taking for granted a vocabulary for phenomenal concepts, we can (seemingly) describe a zombie, entertain the (apparent) possibility that such beings exist, and draw consequences from that (apparent) possibility. But thinking about the acquisition of phenomenal language and concepts brings out latent inconsistencies: here, an inconsistency between the belief that phenomenal language gets its meaning through association with phenomenal experience and the belief that my phenomenal experience is something only I can be aware of. At this point, the thinker cannot retain all of these beliefs. If one does try to hold onto all of them, the result is that we cannot make sense of her words, just as we cannot make sense of someone who affirms p and $\sim p$. One way to describe the status of my own arguments is that, up until the denouement of the second argument, I was participating in an illusion of meaningfulness along with the zombie theorist. The second argument (if it is successful) exposes the illusion. My earlier conclusions are like the conclusions drawn from a belief that is part of an incoherent set: when the beliefs the conclusions rest on are revealed to be incoherent, the conclusions must be relinquished. So also must my earlier argument be relinquished once the zombie hypothesis is revealed to be incoherent.

I will conclude this critique of the zombie theorists’ attempt to define zombies by pointing to a conclusion we should draw from the failure of their attempt. Zombie theorists assume—and in this they follow a major strand of modern philosophy that is explicit in Descartes and implicit in the empiricist tradition—that mental concepts are logically independent of physical ones. Descartes assumed that a thinker’s mind

could be fully stocked with ideas, including the full gamut of phenomenal experience, even if the thinker had no body; this is the possibility the meditator entertains in the Second Meditation. Even once he has established that his body exists, and that his mind is interfused with it so as to be in a certain sense one thing, the meditator argues that the relations between states of his mind and states of his body are purely causal. (Versions of these claims are presupposed by the empiricist argument from analogy.) Against this, Wittgenstein argues that our mental concepts logically presuppose the possibility of the *expression* of those concepts through behavior and action. To have the concept of pain is (in part) to have the ability to express one's pains (through groan or grimace, word or gesture) and the ability to recognize the expression of pain in others. Here, expressing a pain is making it manifest, so that others who possess the concept *pain* can recognize that you instantiate it. And recognizing an expression of pain is perceiving another's pain through her expressing it. The zombie theorist attempts to sever this link between having the concept *pain* and being able to express pain: if zombies are possible, nothing that one does could ever amount to an expression of pain (in the phenomenal sense), since *ex hypothesi* my zombie twin could do the same thing without expressing pain. Furthermore, as we have seen, for a zombie theorist no perceptual intake could ever amount to a perception of another's pain. And what goes for pain goes for any phenomenal concept. In effect, the argument I draw from Wittgenstein about the requirements for giving the words 'felt pain' a meaning is an argument for the logical connection between phenomenal consciousness and the possibility of an outward expression *cum* manifestation of that consciousness. The zombie theorists' denial of any such logical connection leaves them unable to refer to their experience and thereby reduces to nonsense their attempts to propound a philosophical thesis.

Works Cited

- Chalmers, David. *The Conscious Mind*. Cambridge, MA: Harvard University Press, 1996.
- Hyslop, Alec. "Other Minds", *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2018/entries/other-minds/>>.
- Kirk, Robert. "Zombies", *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/spr2019/entries/zombies/>>.
- Marcus, Eric. "Why Zombies Are Inconceivable", *Australasian Journal of Philosophy* 82 (3):477–490 (2004).
- Wittgenstein, Ludwig. *Philosophical Investigations*. 4th ed. Malden, MA: Blackwell, 2009.
- Wittgenstein, Ludwig. *Zettel*. Berkeley, CA: University of California Press, 1967.

WITTGENSTEINIAN (adj.)

Looking at the World from the Viewpoint of
Wittgenstein's Philosophy

Wuppuluri, S.; da Costa, N. (Eds.)

2020, XXII, 560 p. 103 illus., 56 illus. in color., Hardcover

ISBN: 978-3-030-27568-6