

IS CHALMERS'S ZOMBIE ARGUMENT SELF-REFUTING? AND HOW

Julietta Rose

Abstract

In this paper, I will try to show that David Chalmers's Zombie Argument does not show that property dualism about consciousness is true. I argue that the two main premises of the argument are contradictory, and that therefore the argument is self-refuting. The Zombie argument relies on it being the case that I can only conceive of my Zombie Twin if consciousness does not logically supervene on the physical. Chalmers argues that Zombie Twins are coherently conceivable by everybody, if we only try hard enough, and that it therefore follows that consciousness does not logically supervene on the physical. However, I argue that under one conception Zombie Twins are not coherently conceivable, in which case the argument fails; and that under another conception they are coherently conceivable, but show property dualism to be false. In the end, my paper shows that Zombie Twins are not conceivable in the way that Chalmers has argued is necessary, and that they are therefore irrelevant to a discussion of consciousness.

The Zombie Argument as formulated in Chalmers's *The Conscious Mind*¹ is intended to demonstrate that dualism about the mind is true. It employs a two-dimensionalist framework for the analysis of modal concepts and a corresponding logical supervenience relationship² between microphysical facts and more macrolevel facts (such as biological facts, facts about the weather, etc.)³ at the actual world. Both of these positions are in themselves highly controversial; however, I will pass over the objections to be found there, and claim that even if both two-dimensionalism and the logical supervenience relationships that Chalmers postulates are true,⁴ the Zombie Argument fails to demonstrate that dualism is true, and is in fact self-refuting. The argument finds itself in an irresolvable dilemma such that on one horn, it demonstrates that physicalism (re: Lewis) is true, and on the

other, that eliminative materialism, or at least a healthy skepticism about the existence of *any* phenomenal states, is warranted.

Working from the theory that something is primarily (logically) possible if it can be ideally coherently conceived,⁵ Chalmers argues that Zombie Twins—beings who are physically and functionally identical to normal humans, but without conscious experience—are coherently conceivable, and that they are therefore logically, if not actually, possible. It then follows that any description of a conscious agent employing only the physical and functional facts will *actually* underdetermine that agent's phenomenal status; in other words, there is more than the narrowly physical in operation in our phenomenal mental lives, and we need facts beyond the purely physical to describe the phenomenal properties of a conscious agent.

Now, by definition, Zombie Twins must be able to coherently conceive of consciousness, phenomenal concepts, and their *own* Zombie Twins—after all, anyone who “has” a Zombie Twin is doing these very things, and presumably the carrying out of these mental acts is at least partly functionally realized in the psychological mind, not to mention the physical brain. However, the move from “X conceives of her Zombie Twin” to “X's consciousness is does not logically supervene on the physical” requires an account of consciousness under which only conscious agents can conceive of their Zombie Twins with the effect of demonstrating anything about dualism; otherwise, the argument is simply not valid, since X is not necessarily conscious to begin with, so it doesn't follow that X's consciousness fails to supervene. You could say it thus: if some X is conceiving of its Zombie Twin, either (X is conscious, and X's consciousness fails to supervene) or (X is not conscious, and X is somehow failing to refer to the kind of Zombie Twin that would demonstrate dualism).

When Chalmers endorses the above disjunct (or something like it), a very interesting thing happens: Chalmers finds himself denying that consciousness is necessary to conceive of consciousness, form beliefs about consciousness, or construct the Zombie Argument. He instead endorses a deflationary view of conceiving and believing, both for the Zombies and their human counterparts. The problem is that it is unclear what role our consciousness manages to play in *our* phenomenal beliefs and judgments if this deflationary view is correct. Ultimately, the position entails that consciousness is not only functionally inert (which Chalmers is comfortable with), but epistemically inert, so much so that *I* would lack the capacity to

know incorrigibly that I am not a Zombie. If this is so then, given the above disjunct, *my* ability to conceive of *my* Zombie Twin fails to show me anything about consciousness whatever. This result both undermines the argument's viability as a proof of dualism and itself acts as an effective *reductio* against the possibility of Zombie Twins, since it is clearly true that I can know that I am conscious.

It comes down to this: in order for the argument to work, it must be come out that Zombie Twins are physically and functionally identical to me, that they are not conscious, and that I am. However, the framework of Chalmers's Zombie Twins cannot return these results: it seems that, if Zombies are truly not conscious, they must to instantiate mental states such as belief (about phenomenal experience), conceiving (of phenomenal experience), and judgment (about phenomenal experience) in a way that is functionally different from how a conscious thinker would instantiate those "same" mental states. If we want to hold on to Zombies, it means that either *our* capacity to believe, judge, and conceive is just as deflated as the Zombie Twins', and we could be just as mistaken as they are about whether we are conscious; or the Zombie Twins have these capacities in an inflated form identical to our own simply in virtue of their functional identity, they are actually conscious, and physicalism is true.

I will first briefly explain the concepts and terminology necessary to understand the import of the Zombie Argument. I will then present the argument in its most natural formulation, and draw out its intended Dualist conclusions. Finally, I will expose the insoluble dilemma inherent in the Zombie Argument in greater detail and reply to the most potentially attractive strategy for salvaging the argument.

Part I: Fundamentals

Conceivability

The machinery behind the Zombie Argument consists in large part of Chalmers's thesis that the logical possibility of a given proposition⁶ can be determined by its ideal positive conceivability (Chalmers 1999): specifically, a proposition is logically possible when the proposition "is true in some conceivable world," (66) and when "arbitrary details can be filled in with no contradiction revealing itself

on idealized reflection.”⁷ This is distinct from the epistemic sense in which it is conceivable both that Goldbach’s conjecture and its negation (67; also Chalmers 1999) are true. This *apparent* ability to conceive of contradictory propositions is simply due to the fact that we are not ideal conceivers. Thus, we do not have access to the information that would eliminate the false member of the disjunct of “Goldbach’s conjecture is true” or “Goldbach’s conjecture is not true.” Either could be true *for all we know*, but there is no logically possible world in which the false member of the pair is true, or in which both are true. A full a priori positive conception of Goldbach’s Conjecture, with all of the mathematical details filled in, would definitively rule out one possibility. Indeed, it may well be the case that upon further a priori reflection (i.e., intense mathematical investigation) it will become clear that Goldbach’s conjecture *is* true. At that point, it becomes clear that it was never *positively* conceivable that Goldbach’s conjecture was false, and the apparent conceivability of both propositions was based on a *prima facie*, or naive, a priori consideration of the situation. In other words, if Goldbach’s conjecture turned out to be true, there is at this moment no positively conceivable, logically possible world at which it is false—only worlds in which people mistakenly report it to be false.

Primary Intension vs. Secondary Intension

Many objections to the strategy of judging possibility from considerations of conceivability arise due to confusions about the intensions of terms used to denote objects across possible worlds, and the resulting types of possibility that coherent conceptions invoking these intensions are able to entail. Chalmers points out two intensions that almost all terms have, and which can sometimes be confusingly equivocated in cross-world reference: primary intension and secondary intension (57).

The primary intension of a term is the macrolevel manifestation of the referent of the term in some world, considered as actual at the time that reference is fixed. For example, the primary intension of “water” at the actual world, roughly, is “the clear, drinkable liquid that covers most of the earth’s surface, freezes at 32° F, boils at 212° F, etc.”⁸—“watery stuff” for short. As it turns out, “water” as initially characterized by “watery stuff” is identical to H₂O; however, if it had turned out that the primary intension “watery stuff” picked out

XYZ at the actual world, then the term “water” at the actual world would pick out XYZ. There is no a priori entailment from primary intension to secondary intension: people knew about and referred to watery stuff as “water” long before the development of chemistry.

The secondary intension of a term is arrived at after sufficient, usually scientific, investigation reveals the fundamental microphysical properties of the term’s referent at the actual world. At the actual world, the secondary intension of “water” turns out to be H_2O . We can then use the secondary intension, H_2O , to pick out “water” at *any* possible world. Again, it is not logically necessary that the secondary intension of water should match up to the primary intension of water in all possible worlds. H_2O might not necessarily be the “watery stuff” at world X; XYZ might be the “watery stuff,” and H_2O might not exist at all. Thus, if we were using the primary intension to pick out “water” in world X, we would identify XYZ. If we were using the secondary intension, we would conclude that there was none at world X. Additionally, the fact that watery stuff is actually H_2O is necessary only a posteriori and could not have been revealed by any amount of ideal a priori reflection. Without a background knowledge of physics and chemistry and the time to run the appropriate experiments, there is nothing “about” watery stuff that would lead an ideal conceiver to posit that it is identical to H_2O solely on the basis of its being watery stuff.

Thus, the proposition “Water might not have been H_2O ” has different truth values depending on which intension of “water” is invoked. If the secondary intension of water, “ H_2O ,” is invoked, the proposition deals with secondary possibility. It would mean “ H_2O might not have been H_2O .” This statement is clearly false in every possible world, and it follows that the situation expressed in the proposition is not possible. Conversely, if the proposition is interpreted as invoking the primary intension of water, “watery stuff,” it is primarily possible, as there exists a world considered as actual where the watery stuff is not identical with H_2O : the XYZ world.⁹ There is nothing in the concept of “watery stuff” that a priori necessitates that it should be H_2O .

The proposition “ H_2O might not have been water” has a different import. A secondary positive conception would, as above, translate to “ H_2O might not have been H_2O ,” which is clearly false. A primary positive conception, on the other hand, translates to “ H_2O might not have been watery stuff.” This statement could only be true at worlds

with different laws of physics, at which H_2O actually *isn't* watery stuff, arguably due to the fact that the watery stuff isn't H_2O . H and O are defined purely by their causal roles. If these roles are altered, it becomes problematic to maintain that they are still the same kinds of stuff. At the actual world, H_2O could never have been anything other than watery stuff by definition, because at the actual world the properties ascribed to H_2O on the molecular scale directly entail the instantiation of watery properties on the human scale in virtue of the actual laws of physics. Given a completed theory of actual-world physics and a situation in which there is H_2O , it is impossible to conceive of H_2O being other than water, where the term *water* refers to watery stuff. This necessary dependence relation between H_2O and watery stuff is referred to as logical supervenience.

Logical Supervenience and Reductive Explanation

Logical supervenience is a relationship that applies whenever one group of facts or properties (intuitively, the lower-level facts) completely determines another group of facts or properties (the higher-level facts); in a logical supervenience relation, once you have the low-level facts, the high-level facts come along for free. On Chalmers's account, in the actual world a completed physics and a complete microphysical description of the world suffice to entail almost every other fact¹⁰ and property about the world. Biological facts, properties of weather systems and computer systems, not to mention the behavior of the large physical things around us such as water, clocks, and buildings, are all logically supervenient on the microphysical facts. Once a microphysical description is given of cream and sugar chilled to a certain temperature and with certain structural properties, the fact that this stuff is "ice cream" comes along for free.

There is a deep logical connection between reductive explanation, logical supervenience, and primary and secondary intensions. A complete and satisfactory reductive explanation of some phenomenon must, among other things, feature a complete logical entailment from the lower-level explanans (the secondary intension) to the higher-level explanandum (the primary intension). In other words, for a reductive explanation to be successful the properties of the explanandum that are under investigation must be logically supervenient on the explanans; on Chalmers's theory of modal

rationalism, this relationship can be determined by the lack of an ideally conceivable situation in which the two do not co-vary. Reductive explanation succeeds for physical phenomena, and for the things that science manages to explain in general, because the primary intensions of these phenomena are defined by their causal roles. Thus, an explanation of “temperature” as “mean molecular energy” is a complete explanation, because it satisfactorily and without remainder explains all those relations of cause and effect that had been previously labeled as instances of some temperature or another.¹¹ Likewise, an explanation of the capacity to see in the dark will be considered successful if the primary intension of that capacity is defined in terms of behavior, stimuli, and response—finding food at night, avoiding predators and obstacles, etc.—and the explanation provides a coherent and predictive causal account of those capacities. In general, some thing will be amenable to reductive explanation if its primary intension is defined in terms of publicly accessible properties, namely, causes, effects, and interactions or relationships.

The close relationship of ideal positive conceivability to logical supervenience is easily grasped through the following example: Conceive of a bunch of H_2O molecules, at standard temperature and pressure in a situation in which the actual laws of physics hold. Does it follow that you are conceiving of water? While at first it might seem as if you’re not, notice that, given all of the low-level facts already stipulated by the description, there is no alternative reply to the question “What would that H_2O look like to the naked eye?” At the actual world, watery stuff and H_2O are simply the same stuff, considered at lower and higher resolution, respectively: the relationship between the proposition “That stuff is watery” and the proposition “That stuff is H_2O ” is such that the truth value of the former is logically entailed by the truth value of the latter, due to the functional role of H_2O given the explanatory framework of chemistry and physics. It is impossible to ideally positively conceive of H_2O at the actual world without entailing watery stuff; in other words, it is impossible to ideally positively conceive of the secondary intension of water without entailing the primary intension given sufficient background knowledge of chemistry (57–58). It is in this way that we know that the explanatory reduction of “watery stuff” to H_2O is successful. Attempting this kind of thought experiment is one way to determine whether one set of facts or properties is logically supervenient on another set of facts or properties.

Part II: The Zombie Argument

As in the case of watery stuff and H_2O , it would be impossible to ideally positively conceive of Zombie Twins if facts about consciousness logically supervened on the physical facts. The primary intension of the term *consciousness* is roughly, according to Chalmers, the constant experience we all have (assuming the reader is not a zombie) that is characterized by its phenomenal qualities and its essential subjectivity. In this context consciousness does not refer to awareness, of something or in general; the property of being awake; the ability to report on perceptual or mental states; or the property of being self-aware or sentient. Phenomenal consciousness, or phenomenal experience (consciousness and experience for short), is traditionally described as “what it is like” to be a particular conscious agent. What a color, or a bodily sensation, “feels like” is the phenomenal experience of that color or sensation. While it might seem as though this definition itself begs the question, there are many strong reasons independent of the Zombie Argument to accept the claim that this area of experience both exists and is the primary reference of the term *consciousness*.¹² If consciousness is to be explained rather than reduced away, it is precisely this sense of the term that will need explaining.

The Zombie Argument is intended as an a priori demonstration that facts about consciousness do not logically supervene on the physical facts in the way that the properties of the substance denoted by “watery stuff” *do* logically supervene on the physical facts. If the argument succeeds, it will demonstrate that a physical description of a given state of affairs will always radically underdetermine the relevant facts about consciousness; consciousness will never reduce to a secondary intension constituted by physical facts, and thus an explanation of consciousness in terms of physical facts will never succeed. In order to arrive at this conclusion, it is necessary to show that there are two primarily possible situations¹³ such that the physical facts are identical and the phenomenal facts differ: enter my Zombie Twin.

My Zombie Twin is a being who is identical to me as to the physical facts and as to all the properties that logically supervene on those facts, but who entirely lacks phenomenal consciousness. Thus, she is chemically, biologically, and functionally identical to me as well. This is because all of those personal properties straightforwardly

logically supervene on the fundamental physical facts: I cannot conceive of a being who is physically identical to me but who behaves differently; looks different; seeks out different types of food or music; is better at playing soccer; or has a peanut allergy, which I lack. All of these properties are defined in terms of publicly accessible abilities, behaviors, or capacities, which can be analyzed into patterns of cause and effect, and then further reduced down to the arrangement and causal relations between bits of fundamental physical stuff.¹⁴ Once I specify the right arrangement of matter, I get molecular chemistry for free; and once I specify the right arrangement of cells and other physically defined organic stuff, I get a nose, or an eye, or a brain for free. In addition, I get all of the functionality of a nose, or a brain, or an eye. There is no reason to suppose that the causal story of how an electrical impulse moves from the back of my eyeball, down the optic nerve, and into my brain, eventually causing me to exclaim “That cat is green!” will prove to be an enduring mystery. Even if the explanation is incomplete now, it is clearly the kind of thing that we can confidently expect will be explained by a completed psychology/neurobiology. This is precisely because all of the capacities involved in that behavior—exclaim, electrical impulse, optic nerve, etc.—are defined by their causal roles. As such, they are the kinds of public-access conceptual entities that lend themselves to a full explanation in terms of other, more fundamental entities.

So again, consider your Zombie Twin in an analogue of the H₂O thought experiment above. Your Zombie Twin is a being physically, functionally, chemically, and biologically identical to yourself; but this being *entirely lacks* your conscious experience. It behaves exactly like you: prefers the same foods, listens to the same music, expresses the same hopes and desires, is good at the same tasks, holds the same beliefs, etc. However, it is a carbon-based robot. There is as much inner life for your Zombie Twin as there is for a pen, or a fan, or a patch of carpet. There is nothing it is like to be your Zombie Twin (see Nagel 1974).

Got it? Good. Now notice that, by successfully conceiving of your Zombie Twin, you have successfully demonstrated that facts about consciousness do not logically supervene on the physical facts. If facts about consciousness logically supervened on the physical as do the facts about watery stuff, it would be as impossible to conceive of the actual physical facts about yourself holding without the phenomenal

facts as it would be to conceive of the actual physical facts about H₂O without the “watery stuff” facts. The proposition can be stated as:

T1: Consciousness is logically supervenient on the physical iff it is not possible that there exists an X who conceives of its Zombie Twin¹⁵

There is now at least one agent that conceives of its Zombie Twin: you.¹⁶ Therefore, it follows that consciousness is *not* logically supervenient on the physical. More specifically:

T2: It is possible that there exists an X who conceives of its Zombie Twin iff consciousness is not logically supervenient on the physical

The entire Zombie Argument can be stated as:

It is possible that there exists an agent who conceives of its Zombie Twin iff consciousness is not logically supervenient on the physical (T2)

I conceive of my Zombie Twin; therefore, consciousness does not logically supervene on the physical.¹⁷

The failure of consciousness to logically supervene on the physical entails a vast explanatory—and possibly ontological—gap between the properties of structures such as the nervous system and the actual properties of phenomenal experience. If consciousness fails to supervene, there will never be a purely physical explanation of consciousness, on any level of organization, which will logically entail the relevant phenomenal facts. This gap creates a fundamentally unbridgeable divide between consciousness and physical stuff, which lends considerable support to the possibility that some variety of property dualism about the mind is true.

Part III: A Fatal Flaw

Unfortunately, the Zombie Argument has a fatal flaw. In order for the Zombie Argument to go through, my Zombie Twin must manifest every behavior, every brain state, and every mental state that I do. Not only must my Zombie Twin write a paper on the Zombie Argument, she must conceive of her own Zombie Twin. However, it proves impossible for Chalmers to maintain the position that

constructing the Zombie Argument demonstrates the nonlogical supervenience of consciousness, while still delivering on his promise to talk about a concept of “consciousness” that is both coherent and relevant to the actual phenomenon of conscious experience.

As Chalmers himself notes in his discussion of what he terms the “paradox” of phenomenal judgment, my Zombie Twin must also construct the Zombie Argument:

My zombie twin is himself engaged in reasoning just like this. He has been known to lament the fate of *his* zombie twin, who spends all his time worrying about consciousness despite the fact that he has none. ... Still, [my zombie twin] remains utterly confident that consciousness exists and cannot be explained. (180)

From the definition of Zombie Twins, and on Chalmers’s own account, we get T3:

T3: If any X conceives of its Zombie Twin, the Zombie Twin also conceives of its Zombie Twin

However, we can construct two more theses from the definition of Zombie Twins in conjunction with T2:

T4: For any X such that X conceives of its Zombie Twin, X is conscious and its Zombie Twin is not

T5: If some X can conceive of its Zombie Twin, then X is conscious *simpliciter*

T5 is clearly trivially entailed by T4, and the inference from T2 to T4 is definitional. If T2 is true—that is, if consciousness fails to logically supervene on the physical, and if it fails precisely because it is possible that there exists an agent who conceives of its Zombie Twin—then it is true only in virtue of that agent being conscious. As Chalmers himself stresses, only a conscious agent itself can have incorrigible knowledge of its own consciousness, and *incorrigible* knowledge of one’s own consciousness is required for the Zombie Argument to work. Without this knowledge, there is no guarantee that the agent is not a Zombie him or herself; thus, when the agent conceives of its Zombie Twin, there is no guarantee that there are two agents that are identical as to the physical facts but *definitely different* as to conscious experience. A Zombie could no more prove that consciousness fails to logically

supervene by conceiving of its Zombie Twin than I could prove that ectoplasmic ghosts fail to logically supervene: neither of us has the access to the relevant facts (phenomenal facts and ectoplasmic facts, respectively) necessary to see if it makes sense for them to vary independently of the physical facts. The take home message: according to the initial formulation of the Zombie Argument (and some basic intuitions about consciousness) a nonconscious agent can't prove anything substantive about a consciousness that it lacks entirely, let alone that that consciousness fails to logically supervene on the physical.

At this point, the initial dilemma facing the Zombie Argument becomes clear. In order to coherently conceive of my Zombie Twin, I must coherently conceive of her conceiving of her own Zombie Twin (T3). However, by T5 any agent who conceives of its Zombie Twin must be conscious: contradiction. So it seems as if Zombie Twins are either inconceivable, since by definition they cannot instantiate the mental state whereby Zombie Twins are conceived; or if T4 is denied, and Zombie Twins *can* conceive of Zombie Twins, T2 is also thereby denied and the argument crumbles. It follows that if the Zombie Argument works, there is no way we can endorse its central premise T2, since nonconscious agents can conceive of their Zombie Twins as well.

It seems as though, if one doesn't want to forfeit the relationship between the Zombie Argument and consciousness, it's necessary to affirm T5: my Zombie Twin couldn't possibly conceive of her Zombie Twin. Conceiving of a Zombie Twin entails that the conceiver is conscious, but *my* Zombie Twin is by definition *not* conscious; and without conceiving of *her* Zombie Twin, my Zombie Twin is at best a Zombie Counterpart. The proposition "I could have a Zombie Twin" is rendered incoherent, thus primarily inconceivable, and thus impossible. And if it is impossible to conceive of Zombie Twins, then by T1 the Zombie Argument actually proves the logical supervenience of consciousness on the physical.

Part IV: An Unhappy Reply

Chalmers would of course protest this treatment of his argument: he would claim that consciousness on the part of the conceiver is not essential to conceiving of Zombie Twins, only to concluding from their conceivability that consciousness fails to logically supervene on the physical. While the capacity that my Zombie Twin has for conceiving of

her Zombie Twin does nothing to *prove* the logical non-supervenience of consciousness, it also does nothing to invalidate the argument. She is simply instantiating the functional correlate of *my* conception of my Zombie Twin without the phenomenal content, as stipulated.

This reply is the proper one for a defender of the Zombie Argument to make: since on Chalmers's account *everything*¹⁸ about me that isn't phenomenal supervenes logically on the physical, by definition my Zombie Twin instantiates all of my physical and mental properties that are not explicitly phenomenal, including my conception of my Zombie Twin. However, this ploy entails that agents who claim to know incorrigibly that they are conscious on the evidence of their so-called phenomenal experiences could very well be as deluded as their Zombie Twin; in other words, I might be a Zombie. If this is the case, even *my* ability to construct the Zombie Argument fails to prove anything about consciousness, since I might in fact be a Zombie. In addition, it highlights the knotty problem of how Zombie Twins can coherently refer to consciousness while remaining both nonconscious agents and physical/functional replicas of their conscious counterparts.

As we have seen, the claim that only conscious agents can conceive of their Zombie Twins is a direct entailment from T2:

T2: It is possible that there exists an agent who conceives of its Zombie Twin iff consciousness is not logically supervenient on the physical

The problem is that T2 implies that the capacity of a *nonconscious* agent to conceive of its Zombie Twin also proves that consciousness doesn't supervene on the physical, while the credibility of the argument depends upon the opposite. As discussed earlier, a nonconscious agent conceiving of its Zombie Twin—in effect, conceiving of itself—cannot do anything to prove dualism; the possibility that it could is precisely the counterintuitive result of the initial Zombie Argument, which a satisfactory rebuttal must quash.

T2 could be reformulated to stipulate that only conscious agents would be able to demonstrate that consciousness fails to supervene by conceiving of Zombie Twins, while nonconscious agents are just grinding their gears as in the rest of their mental life:

T2': It is possible that there exists a conscious X who can conceive of its Zombie Twin iff consciousness is not logically supervenient on the physical

However, this is simply to *assert* that Zombie Twins can still conceive of their Zombie Twins, and that this does nothing to devalue *our* ability to conceive of Zombie Twins. Paralleling the situation noted earlier, the fallout from this move results in the Zombie Argument refuting itself: there is no way to insist both that Zombie Twins can do everything that we can do, and that what we can do has any relationship to a nonlogically supervenient consciousness.

My Zombie Twin, by definition, instantiates all and only those of my properties that supervene logically on the physical. If Zombie Twins are to be even *prima facie* possible, among my physically supervenient properties must be my belief that I am conscious; my capacity to adopt any number of propositional attitudes toward my supposed conscious experience (I believe that touching a flame will hurt; I like chocolate; red is my favorite color; I believe that my experience of pain proves that I am conscious); and my ability to conceive of my Zombie Twin and draw conclusions about my own consciousness in light of that conception. Any being that did *not* instantiate these mental states and engage in the appropriate behaviors would not be my Zombie Twin.

Is it possible that, in the case of phenomenal beliefs, concepts, and the like, my Zombie Twin has mental states that are functionally identical and yet mysteriously different from my own, and which owe this difference to our differing phenomenal statuses? Is it possible that I have the capacity to discern this difference better than my Zombie Twin, such that *I* can be incorrigible about my consciousness while she is simply mistaken? For Chalmers's position to stand, both of these questions must be answered in the affirmative. Otherwise, I cannot be sure that I am not a Zombie, and thereby incapable of proving anything about consciousness by conceiving of my Zombie Twin. Unfortunately for Chalmers, this is impossible.

If my Zombie Twin is a true Zombie *Twin*, all her mental states must be externally *and* introspectively indistinguishable from my own—I can't think one thing about an experience, while she thinks another. They must occupy the same functional roles in our respective mental lives.¹⁹ Crucially, she wrongly believes that some of these mental states, "experiences," provide incorrigible justification for her belief that she is conscious—just as I believe that my experiences provide incorrigible justification for *my* belief that *I* am conscious. Even according to Chalmers there is no discernible mental difference between me and my Zombie Twin as to our self-ascribed

basis for phenomenal belief. It follows that, at least as far as either of us can *tell*, my beliefs and my Zombie Twin's are equally well justified. How, then, do I gain an epistemic advantage over my Zombie Twin?

Chalmers contends that, due to the fact that I *am* conscious, I have a warrant for regarding my beliefs as true, whereas my Zombie Twin does not, regardless of what we both think about our "experiences." However, how am I to *know* that I am conscious in the first place unless I can *know* that I wouldn't have the same conviction in the absence of consciousness? If my mental life is internally indistinguishable from my Zombie Twin's, then I simply cannot claim to know better about my own consciousness than does my Zombie Twin about her "consciousness." The possibility that I *might* be conscious is totally unhelpful: if Zombie Twins are possible, as far as I can tell I have as much reason and evidence for believing that I am conscious as does my Zombie Twin—hardly a sure thing.

Any further claim that I am somehow warranted over and above my Zombie Twin in regarding my belief that I am conscious as true must invoke some discernable difference between my Zombie Twin and me. And if I can tell the difference, I have at least one mental state that she lacks: I *know that I know* that I am conscious, whereas she does not. If this is the case, then she is not my Zombie Twin, and the argument fails.

To be fair, Chalmers argues that it is my conscious experiences themselves, and not my *beliefs about* those conscious experiences, that form the justification of my claim that I am conscious:

Underlying this sort of objection may be the implicit assumption that the beliefs themselves are the primary determinants of my epistemic situation. ... But of course this is false. The evidence for my beliefs about experiences is much more primitive than the beliefs themselves. It is experience itself that is primary; the beliefs are largely a secondary phenomenon. (199)

In other words, how justified we are in holding certain beliefs is not logically supervenient on the physical—but the beliefs themselves are. We have the same convictions about consciousness as our Zombie Twins, but ours are somehow more justified than theirs. However, the specter of a radically epistemically inert consciousness is not to be gotten rid of.

Chalmers simply pushes the dilemma back a layer: instead of talking about our commonly held beliefs, we are now discussing our commonly held “experiences.” Given that I have the same beliefs as my Zombie Twin, and that we are both convinced that our experiences themselves form the justification for those beliefs, the question again arises: What difference do my experiences make in my mental life that analogous zombie “experiences” don’t make in my Zombie Twin’s? If Zombie Twins exist how could I tell the difference between a real experience and a Zombie “experience” better than my Zombie Twin?

The project of preserving the total functional irrelevance of conscious experience, which is essential to maintain the conceivability of Zombie Twins, reduces consciousness to total epistemic inertia. This strategy maintains the disturbing disconnect between my belief that I am actually experiencing green, and my capacity to *know* that I am experiencing green—there is no available notion of a mentally relevant phenomenal knowledge that I could instantiate and introspectively discern, but which my Zombie Twin could possibly lack.²⁰

Finally, Chalmers claims that the contention that we must have a causally relevant justification for our beliefs in order for them to be “really” justified is a non sequitur: “[F]rom the fact that there is no justification in the physical realm, one might conclude that the physical portion of me (my brain, say) is not justified in its belief. But the question is whether *I* am justified...not whether my brain is” (198).

But the problem extends beyond “the brain.” What Chalmers identifies as the psychological mind—the functionally defined processes that make up the bulk of our phenomenally inert mental life—is a well-established member of the physically supervenient personal properties (as the success of psychology can attest), and beliefs constitute a large part of this psychological mind. If consciousness fails to supervene, we clearly won’t find justification for our phenomenal beliefs in the physical realm; however, it becomes impossible to maintain that we will find discernible phenomenal justification at all. The above position seems to lead into a highly unintuitive (and unproductive) parallelism, at least if Chalmers maintains his interest in preserving the first and second laws of thermodynamics.

The attempt to salvage the Zombie Argument by claiming some vague and unprincipled difference between my Zombie Twins’ “phenomenal” concepts and my phenomenal concepts (no scare

quotes)—a difference that nonetheless has no effect on our respective beliefs about consciousness—results in the crippling indifference of *our* mental life to whether we are actually conscious or not. Apart from being an effective *reductio* against the argument on the face of it, the unprincipled and unwarranted privilege accorded to “actually” conscious agents such as Chalmers and me undercuts Chalmers’s own rationale for why my conception of my Zombie Twin is a viable non-supervenience argument. If T5 is wrong, and any agent can conceive of its Zombie Twin; if only conscious agents can do so with the effect of demonstrating dualism; and if I cannot be sure that I am actually conscious; then by definition, my ability to conceive of my Zombie Twin can demonstrate absolutely nothing about the nature of consciousness.²¹

Part V: The Problem of Reference

The continued defense of the Zombie Argument against the charge of self-refutation results not only in the elimination of a role for consciousness in the formation and justification of our phenomenal beliefs, but in an analogous elimination of our very ability to refer to or even think about actual phenomenal experience. If the Zombie Argument were held to somehow succeed in proving the nonlogical supervenience of consciousness on the physical, it directly follows that the secondary intension of “consciousness” is the same as the primary intension. If consciousness is irreducible, it will not have an *a posteriori* identity with anything but itself. The fundamental structure of consciousness in the actual world would then be nothing except consciousness itself, and *any* vaguely accurate concept of consciousness or vaguely successful reference would derive its meaning and its sense from precisely and only that *actual* consciousness that fails to supervene on the physical. Needless to say, Zombies will have a hard time remaining our twins for long if they entirely lack phenomenal concepts, or if their references for similar sounds diverge functionally from ours—and divergence seems unavoidable given that the terms’ meaning, reference, and sense are all intimately related to nonlogically supervenient consciousness. Strictly speaking, Zombies could not have a concept of consciousness (Moody 1994). It is hard to believe that any agent without such a concept could be my twin in any sense of the word.

Again, it is possible to argue that zombie “phenomenal concepts” are functionally identical to our phenomenal concepts: however, *our* concepts thereby cease to be uniquely ascribable to conscious experience. For Zombie Twins to remain functionally identical—if they are to continue to be possible—the primary and secondary intensions of my concept of consciousness must be shared in common between my Zombie Twin and me; effectively, the primary and secondary intensions must be such that the Zombie Twins have referential access to them, and thus they must *both* supervene on the physical. It follows that the reference and sense of *my* phenomenal concepts, and of my concept of consciousness itself, must also supervene on the physical. But we already said that if Zombie Twins are conceivable, and thus if consciousness does not supervene on the physical, then any concept of consciousness *could not* have been derived from any property or thing or fact in the world which does supervene on the physical:

Nothing in this vast causal story [about the world that logically supervenes on the physical] would lead one who had not experienced [consciousness] directly to believe that there should be any consciousness. The very idea would be unreasonable; almost mystical, perhaps. (102)

The Zombie Argument is again embroiled in an insoluble dilemma: If I have referential access to my own non-supervenient consciousness, then I have a functional property that my Zombie Twin lacks, and Zombie Twins are inconceivable once more; if I want to insist that Zombie Twins are coherent, then I must really reference a property of the world that *does* logically supervene on the physical when I say “consciousness” (I have only as much referential access as does my Zombie Twin). In this case, “real, live consciousness,” if it can even be said to exist, is rendered not only unknowable, but unthinkable and ineffable as such. The Zombie Argument finds itself in the same incoherent position with regard to reference as it does with phenomenal belief: If Zombie Twins are possible, then the argument is false or meaningless, since any reference to consciousness will pick out something that supervenes on the physical; if the argument is sound, then Zombie Twins cannot be functionally identical as regards the references of their phenomenal terms, and they are inconceivable.

Unlike almost all other mental entities or properties, the concept of consciousness itself seems to be unrealizable in mental states that lack the appropriate contact with the phenomenal realm. Chalmers himself admits this:

Somehow a sort of experience, which one might think of as the *referent* of a qualitative concept, is getting inside the [phenomenal] concept and constituting its sense (where sense is equated with primary intension).” (207)

The term *consciousness* could have no other referent and no other sense than real, live, phenomenal consciousness. If its reference, meaning, primary and secondary intensions, and sense are all to be found in nonlogically supervenient consciousness, it follows that functional role is must be derived entirely from the nonlogically supervenient consciousness; and it follows that zombies *cannot have* a functionally identical concept of consciousness—not even an “empty” one. An empty concept of consciousness would not fill the same causal role in our mental life, speech acts, or behavior, and it is unclear that there would even be a “container concept” available to be filled.

Chalmers claims, as it seems he must, that the Zombie Twins possess some limited concept of consciousness—perhaps as a “property over and above any physical and functional property” (204). But Chalmers also characterizes consciousness as “something surprising” about the world, something which no amount of a priori or a posteriori investigation into the physically supervenient stuff in the world would lead one to expect—and so he should. This is the position on which the entire Zombie Argument is predicated. What possibility, therefore, exists that a being living in a world where *everything* supervenes logically on the physical would posit something that lies over and above it, and which forms the *ultimate content* of their purely physically supervenient and functionally defined “phenomenal” concepts and beliefs—concepts and beliefs that *have no content*? To assert that the Zombies will have any notion of consciousness at all, while maintaining that the *only* meaning consciousness can have in statements such as “I believe I am conscious because I have the experience of pain” is *phenomenal, nonfunctionally defined experience*,²² is simply contradictory.

Chalmers wants to have it both ways: he wants *us* to be able to refer to consciousness thanks to the brute fact of our phenomenal experiences; and he wants Zombies to refer, inaccurately, to their “consciousness” thanks to their quasi-experiences; but he does not provide any rationale for how the Zombies could ever refer to anything that would be patently absent from their scope of reference or thought.

The possession of such a concept by our Zombie Twins, even one that only *functioned* like our concept, would undermine the necessarily and essentially unique position occupied by our phenomenal concepts and phenomenal beliefs. Strictly speaking, the Zombies Twins could not have the concept; to insist that they would, even if only a functional analogue, opens the door to the possibility that the referent of *our* term *consciousness* logically supervenes on the physical.

In other words, to maintain both the conceivability of Zombie Twins and the validity of the Zombie Argument entails two contradictory conclusions: first, that consciousness entirely fails to supervene on the physical, and second, that Zombies can competently use the concept of consciousness in their thinking and speech acts. If they can reference consciousness without any kind of connection to consciousness, then we can too—and what’s more, we are as bad at telling the difference as our Zombie Twins. It follows that the Zombie Argument is simply nonsensical, since what we *actually* mean by consciousness logically supervenes on the physical. Any concept of consciousness that would be worthwhile—a concept that would not be accessible to nonconscious agents—would be not only unknowable, but entirely inaccessible to thought or reference. Unless we are again given an unprincipled privilege over our Zombie Twins, the coherency of a Zombie Twin entails that it is impossible for us to ever refer to actual consciousness as defined by the Zombie Argument: whatever we mean when we say “consciousness,” it definitionally *does* logically supervene on the physical.

Part VI: Finale

David Chalmers’s Zombie Argument is probably the most influential defense of dualism in contemporary literature. Nonetheless, the argument suffers from a crippling, hidden self-contradiction: if conceiving of Zombie Twins is restricted to conscious agents, as

seems *prima facie* necessary for the argument to go through, then the Zombie Twins themselves are rendered inconceivable; but if any agent with the requisite cognitive capacity can conceive of its Zombie Twin, then my own status as a conscious agent is called into question, and my ability to conceive of my Zombie Twin can be chalked up to the live possibility that I am in fact a Zombie. In neither case does my conception of my Zombie Twin do anything to demonstrate dualism—in fact, the opposite is true. In the first case, consciousness is seen to clearly supervene on the physical, at least in the sense necessary to imply the possibility of a reductive identification of consciousness with some physically supervenient aspect of the mind. In the second case, phenomenal consciousness is simply eliminated entirely from mental life, even on introspection. It is simply a phantom, a quasi-concept with no real meaning or possible referent. It turns out that if the conceivability of my Zombie Twin can be used to argue for dualism, then she is inconceivable, and if she is conceivable, then her conceivability cannot be used to argue for dualism.

Further, the argument threatens Chalmers's most basic tenet: thou shalt not say "consciousness" and mean something else (26). If the Zombie Argument went through, it would follow that any reference I might make to consciousness, phenomenal experience, qualitative states, and the like would surreptitiously pick out a functional property of my mind that I share with my Zombie Twin. In the process, the argument would either prove that consciousness is beyond the limits of meaningful thought and reference and that the argument itself is thereby moot; or that agents that are stipulated to lack consciousness have as much access to it as do I.

Taken together, these considerations support an interpretation under which the Zombie Argument forms an impressively strong *reductio ad absurdum* argument in support of some form of eliminative reductionism about phenomenal experience, at least if the strain of modal rationalism that endorses propositions such as T1 is taken seriously. When you run the thought experiment, the dualist account of consciousness breaks down under the weight of its own theoretical commitments.

However, it is my position that regardless of the failure of the argument, eliminativism about qualia is clearly false: I am clearly having conscious experiences, I am clearly able to think about them and refer to them, and they clearly form the justification for my belief that I am conscious and that I see colors and feel sensations. What's

more, they form a stronger justification than do their neural or functional mental correlates. To endorse any position to the contrary would be an exercise in self-delusion. As John Searle points out in *The Mystery of Consciousness*, reduction that involves wholly eliminating the explanandum in favor of some other set of facts or properties is based on a distinction between (illusory) appearance and reality (Searle 1990, 29): rainbows, initially conceived of as solid ribbons of color arching across the sky, can be eliminated in favor of a story about the refraction of light through water molecules present in the atmosphere after it rains, because *there are no* solid ribbons of color arching across the sky. However, consciousness cannot be fit into the illusion/reality dichotomy; conscious experience is the center of our epistemic worldview, the incorrigible base from which we venture out into the possibility of the world beyond experience. If conscious experience is rendered illusory and untrustworthy, there is simply nothing left. Further, if there is the “illusion” of conscious experience, the experience of that illusion just *is* conscious experience, Zombie “consciousness” notwithstanding.

Rather, it seems that there is some deep connection between human (and probably most animal) minds, functionally characterized, and consciousness itself—you cannot have one without the other, even when it seems like you can. On the inductive evidence available from the scientific community today, there is no reason to suppose that phenomenal consciousness will ever form part of the reductive explanation for a behavior, brain state, or even a mental state functionally defined, or itself be reductively explained; it seems entirely nonessential to the whole enterprise. But on internal reflection, actual consciousness *is* essential to our ability to form concepts that characterize it, terms that reference it, and beliefs that endorse (or deny) its existence.

Unless conceivability is abandoned as a guide to possibility entirely, it seems that the only recourse for a conscious essentialist—to use Flanagan’s term (Moody 1994 from Flanagan 1991)—is to reaffirm T5: in order to play the game of talking about consciousness, you must be conscious yourself. A nonconscious agent, although it could certainly learn to use words such as “consciousness,” “phenomenal experience,” and the like, would never originate them if left to its own devices (also see Moody 1994). Further, nonconscious agents cannot make sense of using these words to the same degree as would a conscious agent, since their scope of reference does not extend to

the phenomenal realm; it follows that conscious agents and Zombie Twins, if they make all the same sounds, must have two homophonic, but distinct, vocabularies.

What follows from this? Does consciousness, as it is experienced prior to conceptualization, actually logically supervene on the physical? Is it actually reducible, without remainder, to some physical or functional correlate? If this is so, the deep-seated intuitions informing the dualist view that consciousness is nonreducible, such as those that make Mary, Zombies, and Inverted Qualia seem plausible, must be explained away by a new conception of consciousness that does not lead to an underdetermination of phenomenal states when employed within an appropriate explanatory framework. In addition, these intuitions must be rendered toothless in a way that makes it intelligible *why* they were so compelling in the first place.

Some reductionists about consciousness suggest that with more empirical data, the sense of “consciousness” that needs explaining will be modified into a concept which is defined by its causal role alone (such as Block and Stalnaker 1999). That is, just as progress in the biological sciences showed us that “to be alive” can be defined as a collection of functional characteristics of some systems, and thus that the question of whether something is alive or not can be settled by noting certain of its physical or functional properties, what we think it is “to be conscious” will prove similarly amenable to scientific “therapy.” Eventually, we will regard questions such as “Is Chalmers *really* conscious?” on par with questions such as “Is the life force a physical property?”

A redefinition of consciousness in this way faces obstacles on a different scale than the analogous redefinition of life (108–109). Many take the position that this project is impossible, due to what we already know about the nature of consciousness itself from our own mental lives (Dietrich and Hardcastle 2004). Others feel that this possibility cannot be ruled out, and the attraction that they feel for this kind of “wait until next year” reply is understandable. The failure of the Zombie Argument may be seen to provide ammunition against any conception of the mind that draws some kind of in-principle insurmountable distinction between consciousness and other aspects of the mind. However, it seems to me that the failure of the Zombie Argument is a surprise, and instead serves to prove how little we understand about even our naive conception of consciousness. Any reduction or identification of consciousness to purely physical or functional properties of mind would simply be sweeping the

problem under the rug—it's true, we wouldn't have to look at it or think about it, but the weirdness would still be there.

The weirdness of consciousness seems to originate, as suggested by Nagel (1974), in its essentially and uniquely subjective nature. We are helpless to show our phenomenal experiences to anyone else, but as Chalmers points out, consciousness constitutes an essential component of our core epistemic situation (195). If the Zombie Argument demonstrates anything, it is that the epistemic boundary between the minds of separate conscious agents is so high, and so insurmountable, that it makes the idea that *my* mind could exist without consciousness—without *me*—seem positively conceivable. The question now, which may be unanswerable in principle, is *why*.

NOTES

1. All further unspecified citations refer to Chalmers (1996).
2. Logical supervenience is a relationship between two sets of facts (A-facts, the lower-level facts, and B-facts, the higher-level facts) such that it is logically impossible for the A-facts to vary without the B-facts to co-vary. The A-facts form the supervenience base for the B-facts (Chalmers and Jackson 2001).
3. In addition to two other minor classes of facts. In Chalmers and Jackson (2001), they argue that almost all types of facts are logically supervenient on the physical—i.e., they can be derived a priori from a complete description of the world in terms of physics. Phenomenal facts, limiting or “that’s all” facts, and indexical or centering facts do not logically supervene on the physical facts. He calls this scheme PQTI: Physical, Qualitative (phenomenal), “that’s all,” and indexical, respectively.

All further references to “microphysical facts,” “fundamental physical facts,” etc. should be understood as a statement of the conjunction of all of the facts that are entailed by P, T & I. Obviously, whether this includes the Q facts or not is dependent on the soundness of the Zombie Argument.

4. Perhaps, especially if they are true.
5. (Chalmers 2002). Ideal conceivability, on Chalmers's view, simply means that there are no logical defeaters of the position on ideal reflection. One interesting corollary of this view is that a proposition/situation/state of affairs that is as of yet not “inconceivable” always has the epistemic possibility of being defeated on further reflection. The only possibilities of which we can be sure about their conceivability are those that have been shown to be inconceivable.

6. A proposition can consist of the conjunction of any number of basic propositions about potential states of affairs, such as “Grass exists” or “I live in Binghamton.” The statement “Grass exists and I live in Binghamton” is also considered one proposition—it is one way the world could be.
7. A priori reflection carried out by an ideal rational agent. This is distinguished from *prima facie* reflection: more superficial reflection, which may fail to expose inconsistencies or contradictions inherent in the proposition (Chalmers 1999).
8. Some of these attributes—such as the exact boiling and freezing points—are obvious only after considerable empirical investigation. Many, such as potability, freezing and boiling, etc. are dependent on considerations that go beyond whether some stuff is H_2O —sea water has a slightly different freezing point, and is not especially drinkable. However, these discrepancies are also explained by the same physical theories that make an explanatory reduction of watery stuff to H_2O successful and convincing. The picture presented here is a necessary simplification of the actual conceptual territory, but this simplification shouldn't have any effect on the viability of the argument.
9. It is important to note that Chalmers does not advocate the position that the ideal primary positive conceivability of the secondary possibility of some proposition entails the actual secondary possibility of that proposition. For example, it is primarily positively conceivable that it is secondarily possible that water might not have been H_2O (i.e., that water is not H_2O at the actual world). There is no a priori necessity in the identity at the actual world of watery stuff and H_2O . However, it remains the case that water is H_2O . Thus, primary conceivability can never serve to fix secondary possibility, the state of affairs at the actual world. Secondary possibility is fixed a posteriori, and since it turns out that water is H_2O , it is not secondarily possible that water is not H_2O . However, this is not a problem for the Zombie Argument, as the structure of the argument only requires the primary possibility of Zombie Twins, which is fixed a priori (Chalmers 1999, 197).
10. Such things as causality, moral facts, mathematical facts, and facts involving vague predicates might not be supervenient on the physical facts alone (Chalmers 1999). Also, as mentioned before, limiting facts and indexical facts are not held to be supervenient on the physical facts (Chalmers and Jackson 2000).
11. Assuming that the promise to cash out the terms of the explanation can be fulfilled.
12. The foremost of these reasons is the simple brute fact of conscious experience. In addition, the Mary Argument (Jackson 1986), the Inverted Qualia Argument, the Problem of Other Minds (the Zombie Intuition), and the Cartesian Intuition lend support to a view of phenomenal consciousness on which it is intrinsically and essentially subjective. In addition, the acknowledged failures

of behaviorism, physicalism, functionalism, and other materialist theories of mind to precisely capture this aspect of the mind, and the continued failure of psychology, neurobiology, etc. to even invoke consciousness in explanations of such mental phenomena as color vision, point to the conclusions that (1) methods of describing and explaining the mind that approach the problem using third-person accessible explanatory data and causally defined “mental mechanisms” as their building blocks will not have anything to say about consciousness and (2) this inability to address the problem of consciousness head-on stems directly from these third-person methods of inquiry and explanation.

13. These situations only have to be primarily, or conceptually, possible. Whether or not zombies are actually possible at the actual world is irrelevant to the question of logical supervenience.
14. Ignoring, for the moment, the reduction of bits of physical stuff to physically defined waves, points of energy, etc. This detail plays no material role for the argument.
15. If humans are conscious, then dogs are clearly conscious as well. However, they aren't smart enough even to grasp the concept of Zombie Twins, let alone conceive of them. Nevertheless, if only dogs populated the planet, the relationship expressed in T1 would still be true, because it would still be possible in principle for there to exist a conceiver with the requisite level of intelligence to attempt to conceive of Zombies. However, it's unclear whether this possibility can be known a priori, or whether it only obtains because humans actually happen to exist. In any case, they do exist, and they do attempt to conceive of Zombies.
16. If you find yourself unable to do the thought experiment, it might be that you are misconceiving (96). On the other hand, as Daniel Dennett has claimed in his own case, you simply might not be able to square your deepest materialist intuitions with a robust conception of your Zombie Twin (Dennett 2005). However, assume for the sake of argument that you can, if only you try hard enough—after all, Chalmers can. A simple denial of one's ability to conceive of Zombie Twins is not sufficient to defang the argument.
17. The conclusion follows in virtue of the meaning of the concept of consciousness Chalmers is employing: since consciousness is an inherently private-access property, the only intension I could initially have for the term *consciousness* is the one I derive from my conscious experience. If the initial construction of the Zombie Argument is successful and my consciousness actually fails to logically supervene on the physical, then it follows that there is simply nothing else in the outside world from which I could have acquired the concept. Thus, if I choose to ascribe consciousness to another entity, the only property that I could be ascribing is consciousness as I understand it, which supposedly does not logically supervene on the physical. A successful execution of the Zombie Argument would show that all instances of consciousness fail to supervene, period.

18. Excepting, of course, indexical facts and a “that’s all” fact.
19. Similar argumentation can be found in Nigel (1998).
20. This strategy should be distinguished from one discussed on pp. 192–93. The contended indiscernability of my and my Zombie Twin’s conscious experiences, and our beliefs about those experiences, is not due to their having been formed through the same mechanisms. Rather, it is a matter of internal indiscernability: on introspection, I have no way to discern whether my real-seeming experiences are actually real, over and above the way that they seem real to a Zombie. Ironically, Chalmers makes the same mistake that he accuses Dennett of making on page 190: he confuses introspection with extrospection.
21. Again, Nigel (1998) makes a similar point.
22. This is the meaning it must have in order for it to be possible to coherently challenge physicalist and functionalist theories of mind.

WORKS CITED

- Block, N., and R. Stalnaker. 1999. Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review* 108, no. 1: 1–46.
- Chalmers, D. J. 1996. *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.
- . 2002. Does conceivability entail possibility? In *Conceivability and possibility*, ed. T. S. Gendler and J. Hawthorne. Oxford: Oxford University Press.
- , and F. Jackson. 2001. Conceptual analysis and reductive explanation. *Philosophical Review* 110, no. 3: 315–61.
- Dennett, D. C. 2005. *Sweet dreams: Philosophical obstacles to a science of consciousness*. Cambridge: MIT Press.
- Dietrich, E., and V. G. Hardcastle. 2004. *Sisyphus’s boulder: Consciousness and the limits of the knowable*. Amsterdam: John Benjamins.
- Flanagan, O. J. 1984. *The science of the mind*. Cambridge: MIT Press.
- Jackson, F. 1986. What Mary didn’t know. *Journal of Philosophy* 83 (May): 291–95.
- Levine, J. M. 1983. Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64 (October): 354–61.

- Moody, T. C. 1994. Conversations with Zombies. *Journal of Consciousness Studies* 1, no. 2: 196–200.
- Nagel, T. 1974. What is it like to be a bat? *Philosophical Review* 83 (October): 435–50.
- Searle, J. 1990. *The mystery of consciousness*. New York: New York Review Books.
- Thomas, N. J. T. 1998. Zombie killer. In *Toward a science of consciousness II*, ed. S. R. Hameroff, A. W. Kaszniak, and A. C. Scott. Cambridge: MIT Press.