

Dietrich, E. (1998). Zombies only seem logically possible: or How consciousness hides the truth of Materialism: A Critical Review of The Consciousness Mind by David Chalmers. *Minds and Machines*, 8 (3), pp. 441-461.

**A Review of *The Consciousness Mind***  
by David Chalmers, Oxford Univ. Press, 1996.

Eric Dietrich  
Philosophy Dept.  
Binghamton Univ., Binghamton NY 13902

## 1. Introduction

When Charles Darwin died in April, 1882, he left behind a world changed forever. Because of his writings, most notably, of course, *The Origin of Species*, by 1882, evolution was an almost universally acknowledged fact. What remained in dispute, however, was *how* evolution occurred. So because of Darwin's work, everyone accepted that new species emerge over time, yet few agreed with him that it was natural selection that powered the change, as Darwin hypothesized. Chalmers' book, *The Conscious Mind*, reminds me of *The Origin of Species*. I have talked to many people about *The Conscious Mind* and watched many philosophy students read it, and after they are done, they all take consciousness more seriously than they did before and they are all struck by its deep mystery, yet few accept Chalmers' specific theory of consciousness and his explanation of its odd nature.

Chalmers has written an exciting and fascinating book. And I for one hope that because of it, consciousness in all its paradoxical glory will once more hold center stage in a robust philosophy of mind and metaphysics. The book is good enough that this is a reasonable hope. However, I fear that the analogy with Darwin's book will not be sustained. The theory of natural selection unveiled in *The Origin of Species* was eventually vindicated, and nowadays, we regard Darwin's theory of natural selection as largely correct. But I doubt that the same fate awaits Chalmers' positive theory of consciousness. So while Chalmers' book will be a source of great philosophy and provoke dozens of papers, I think at the end of the day his contribution will be to have shown us the problem of consciousness in all its profundity, yet consciousness will remain an unsolved problem.

Comparing Chalmers' book with Darwin's yields two other points of interest. Chalmers, like Darwin, has shown us a world that is richer, stranger, and more complicated than what we thought. This is true even if Chalmers' theory is not correct because he has made the problem of

satisfactorily explaining consciousness so pointed. There are philosophers and other cognitive scientists who think that consciousness doesn't even exist; there are those who think it exists, but reduces to something else such as neurocomputation or metaprocessing or the quantum states of microtubules; and there are those who, like myself, hoped consciousness would reduce to something, but were unsatisfied with all current proposals. For all three groups, Chalmers has complicated and unsettled the world. Consciousness is real but it cannot be reduced to anything physical, not neurocomputation, nor metaprocessing, nor even quantum mechanics. Chalmers, in short, is a dualist. According to him, consciousness is *sui generis* and completely unlike any physical process. And here we see the great disanalogy between Chalmers' and Darwin's books. Whereas Darwin explained how some strange and complex phenomenon (evolution) could be a purely mechanical process (natural selection), Chalmers has done just the opposite. He has shown how a strange phenomenon can *never* be a mechanical process.

In this review, I will explain how Chalmers arrives at the conclusion that materialism must be false, I will explain his positive theory of consciousness (that a sort of dualism must be true), and finally I will discuss an objection to his theory. Doing all of this still leaves out much of the book. It is rich and long (about 395 pages, including notes, which can be skipped by non-philosophers, but only non-philosophers). Still, doing this much will give the gist of Chalmers' book, and, I hope, excite the reader to spend some time with it and then ponder the nature of consciousness.

## 2. Groundwork

Before we delve into his anti-materialist arguments, we need to spend a little time getting clear on what the topic actually is. This Chalmers does in the introduction and chapter one.

What is consciousness? Consciousness is the way the world seems to us, the way we experience it, feel it. Bite into an onion, look at a rainbow, sniff a dead skunk on a hot summer's day, stub your toe on the foot of the bed frame at four A.M., listen to a baby gurgle and coo. These are experiences, bits of our phenomenology or qualitative feel of the world (philosophers sometimes use the term *qualia* to refer to these qualitative feels), and it is experiences, qualia, that give us our subjective point of view. We have experiences because we are conscious. Or rather, our having them *constitutes* our being conscious (can you be conscious without being conscious *of* something?) Being conscious is what makes it fun or horrible or merely boring to be a human. Using the phrase Nagel made famous (1974), Chalmers says "We can say that a being is consciousness if there is *something it is like* to be that being ..." (p. 4) [ftnt1].

Conscious experience is the most familiar thing in the world. You know nothing as certainly as your own conscious experiences (the contents of your own conscious states), and next to that, nothing as certainly as the fact that you are conscious. This is quite odd, Chalmers notes, given that the inexorable march of science has somehow managed to say nothing illuminating about consciousness. This may be one of the many spots where philosophers and others will disagree with Chalmers, but the statement is true nevertheless. If *science*, not just a few individual scientists here and there, had managed to say something substantial about consciousness, then there would now be some agreement as to what consciousness is and how to explain it. Compare the evolution of life, or the behavior of masses in a gravitational field. Science has said something illuminating about these phenomena, and that is just why there is agreement about them. But there is absolutely no agreement on consciousness, neither what it is, nor how to explain it. So, oddly, the thing we as individuals know best, our conscious experience, is the thing about which we as a collective of understanders know least.

This is a crucial point, so I want to stress it a bit. Chalmers' whole theory begins here: consciousness is a big, deep mystery. It is completely surprising, both that it exists and that it exists the way that it does. Why hasn't science had anything useful to say about consciousness? It seems that a complete and detailed catalog of the physical, chemical, biological, and computational facts of the world would not in any way entail, suggest, or even hint at the existence of consciousness. Yet there it is in all its glory. Consciousness is utterly strange.

The two fundamental questions about consciousness that we can't answer are 1) why is it like something to be a human, or any other consciousness being, and 2) why do we have the particular experiences we have? Why is like anything at all to bite into an onion, and why does biting into an onion lead to the particular experience that it does? Why don't we experience blue when we bite into an onion? These are very difficult questions, and, as Chalmers points out, we really have no idea how to answer them. His book is devoted to laying the groundwork for answering these questions.

Chalmers exercises our intuitions about consciousness first by presenting a small catalog of conscious experiences ranging from visual experiences, through mental imagery, to emotions, to our sense of self, and then by drawing a crucial distinction between consciousness, on the one hand, and the rest of our psychological mind, on the other. He calls these two the *phenomenal* concept of mind and the *psychological* concept of mind. The phenomenal concept (or aspect) of mind considers mind as the locus of conscious experience. The psychological concept considers

mind as the causal and explanatory basis of our behavior. For example, neural responses to wavelengths of light between 630 and 750 nm beginning with cones in your retina and ending somewhere back in your visual cortex can lead you to believe that you are looking at a red object. The perceptual processes leading to the belief as well as the belief itself are psychological aspects of mind. They are studied by psychology (I will return to this shortly). It is true that you also *experience* red as you look at the object, but, at least *prima facie*, your experience of red is different from the perceptual and cognitive processes that underlie it. Chalmers is careful not to beg any questions here. The distinction between phenomenal and psychological aspects of the mind is an “in-principle” distinction -- a distinction based on meaning. These two aspects of mind are not related in any sort of obvious conceptual way. And that is all Chalmers needs to get started. The two might turn out to coincide intimately (so intimately as to be identical, like the morning star and the evening star), but they might not, too. Only inquiry will tell.

Other examples of the psychological concept of mind include learning, memory, motor control, and unconscious perceptual responses. Of course, all unconscious mental states are examples of the psychological aspect of mind, since they don't even have a phenomenal component.

Above, when I said that science has told us little that is illuminating about conscious experience, I did not mean to imply that science has told us nothing about the mind. It has. But what it has told us has little to do with our consciousness, rather it has to do with how the mind operates, how it responds to inputs and causes behavior. The science of the mind -- psychology in the large (including all of cognitive science and artificial intelligence as well as clinical psychology, evolutionary psychology, brain science, and others) -- studies the psychological conception of mind almost exclusively, eschewing the phenomenal conception. Virtually everything we know about the mind concerns its role as a *device* that responds to the world and produces behavior.

Chalmers further motivates the distinction between the phenomenal concept or aspect of mind and the psychological aspect with this heuristic: “A good test for whether a mental notion M is primarily psychological is to ask oneself: Could something be an instance of M without any particular associated phenomenal quality?” (p. 18). If the answer is “Yes,” then you are probably dealing with a psychological notion, and one that the science of psychology (in the large) can study. If the answer is “No,” then you are probably dealing with a phenomenal notion. Of course, many of our psychological states have a phenomenal component and vice versa, but the two notions of mind are nevertheless distinguishable at least in principle (like in the perception of the red object, above). As I said, the phenomenal and psychological aspects of mind might turn out

to be the same thing. In fact, this is what the materialist hopes for. But they might also turn out to be different things, or different properties of the same thing. This is what the dualist hopes for. And what Chalmers argues for.

Finally, there are three other matters to discuss before we can turn to Chalmers' arguments. First, Chalmers insists on taking consciousness -- the phenomenal aspect of mind -- seriously. Taking consciousness seriously simply entails not conflating it with the psychological aspect. We are going to have to have a good argument before we identify them with each other (and, of course, we will have to have a good argument before we say they are completely different). Chalmers briefly reviews the history of these two conceptions of mind, showing how first, philosophers identified the psychological with the phenomenal, making the phenomenal all-important, and eventually ushering in introspectionism (this arguably began with Descartes), and second how in this century, many philosophers and psychologists reversed this trend and tried to identify the phenomenal with the psychological. This didn't work either. The bottom line is that both the phenomenal and the psychological are important aspects of mind and neither should be thought to be whole story.

The second matter is that Chalmers approaches dualism and the study of consciousness in general from a completely nonspiritual, nonreligious perspective. Though he doesn't play this up, I found it to be one of the most refreshing aspects of the book. Here is a dualist who does not in any way fear the mechanistic forces of darkness (Dietrich, 1995?). In a sense, he wants to be a "mechanistic" dualist; the dualism he advocates is governed by natural laws, it is just that these natural laws extend beyond the laws of our current sciences. Indeed, he calls his theory *naturalistic dualism*. Chalmers is an ardent supporter of artificial intelligence and the computational theory of mind. He might even be unique among dualists in believing that the psychological aspect of mind can be completely explained within the computational paradigm. This is important because it is what makes the book compelling. Chalmers manages to come across as a dualist who is not a mysterian. It is via this that he shows us a world that is richer and more complicated than we have imagined. Mysterians are easy to dismiss because they are really erecting Maginot lines around the human psyche, trying to protect it from the onslaught of science. Chalmers will have none of this. He wants to *scientifically explain* the human psyche -- it is just that we are going to have to extend our science if we are going to be successful.

The third matter is arguably the most crucial. Chalmers suggests in chapter 1 (and says explicitly in chapter 5, p.186) that though consciousness is surprising, claims or reports about consciousness are not. Chalmers has to have this, as we will see, but this move gets him into a

world of grief later in the book; we will return to it below. For now, let's merely note that Chalmers draws the distinction between the phenomenal and the psychological so tightly that verbal reports of (alleged) phenomenal states, which are in fact due to psychological states, are regarded by Chalmers as not mysterious, whereas consciousness is the supreme mystery.

### 3. The Argument against Materialism

Chalmers' arguments against materialism and for dualism are unusual in that he does use the notion of identity, e.g., psychophysical identity is never used at all. But even though he doesn't invoke identity, Chalmers still uses possible worlds. This is another place where he is vulnerable to attack. If you are dubious of possible worlds (and well you might be) then many of Chalmers' arguments will probably also strike you as dubious. I will discuss this below, in section 6.

Instead of identity, Chalmers uses the notion of *supervenience* (Davidson, 1970; Kim, 1978, 1993). This not only works rather well, but is refreshing and allows him to do some nice metaphysics. Chalmers introduces the technical notion of supervenience and discusses its relation to explanation in chapter two -- a long and philosophically technical chapter, but a rewarding chapter, too. The key notion of supervenience is *logical supervenience*. This is the only notion I will be concerned with here. The definition of logical supervenience is as follows (p. 33):

B facts/properties logically supervene on A facts if no two logically possible worlds are identical with respect to their A facts while differing in their B facts.

With this distinction in hand, let's move on to Chalmers' arguments. (Part of what happens in chapter 2 is that Chalmers shows how supervenience and reductive explanation are related. This is a nice discussion, but looking at it would take us too far afield. I am just going to summarize this by saying that a physical phenomenon is reductively explainable in terms of some low-level physical properties if and only if it logically supervenes on those properties (pp. 47-51).)

The arguments against materialism begin with the intuitive distinction, introduced above, between the psychological aspects of mind and the phenomenal aspects of mind. I noted there that though we might discover that these two aspects name the same thing, they are nevertheless conceptually different -- different because of the *meanings* of the terms, just like the terms "morning star" and "evening star" which both pick out Venus. Recall that above I also

introduced Chalmers's heuristic for distinguishing between the psychological and the phenomenal aspects of mind: "A good test for whether a mental notion M is primarily psychological is to ask oneself: Could something be an instance of M without any particular associated phenomenal quality?" Chalmers key insight is that the mere existence of this heuristic guarantees that materialism is false. The fact is, Chalmers argues, that we can imagine *all* psychological aspects of mind occurring without any particular associated phenomenal qualities. And this fact means that phenomenal aspect is not a part of the physical world the way everything else is. Here is how the argument goes.

Everything in the world logically supervenes on the level below it. Fix the low-level physical facts of our world, the behaviors and trajectories of every particle -- every quark, electron, proton and neutron -- and you automatically fix all the other facts in our world -- the chemical facts, the biological facts, the psychological facts, and the social and cultural facts. In other words, it is logically impossible to imagine a world just like ours at the lowest level, that has exactly the same detailed, low-level physical facts as our actual world has, but which differs from our world in its high-level facts.

Here's an example using a glass of water. Imagine a glass filled with hot water. The atoms in the glass are caroming all over the place in a very agitated way. Now, try to imagine another glass of water where the atoms are behaving in *exactly* the same way as in the first glass, but where the water in the second glass is cold. You can't do it. Or, if you think you can, you are mistaken (c.f, p. 109, top). For, all we *mean* by "hot" is that the atoms are caroming all over the place in a very agitated way. Fix the behavior of the water atoms in the glass and you automatically fix the water's temperature. This example exhibits just what is going on at the level of our entire universe. It is simply inconceivable that the low-level facts about our world could be what they are and yet there be no stardust, no suns, no galaxies, no planets, no continents, no minerals, no life, no US Constitution, no penguins in Antarctica, and no MTV (the Music Television Channel). In short, and though it may sound strange, MTV logically supervenes on the low-level physical facts of our world. There is no possible world with the same low-level facts as ours that isn't blessed with MTV. This supervenience hierarchy subsumes everything; everything in our world supervenes *logically* on the level below it and ultimately on the lowest level -- everything, that is, but consciousness.[ftnt2]

How do we know that consciousness doesn't logically supervene? Because we can imagine *zombies* (This is chapter three. Chalmers actually gives five arguments for this conclusion, but the zombie argument is perhaps the most compelling; I will discuss all five shortly). Zombies are

creatures who behave exactly like us, they run around, they laugh, they cry, they dance to music, and yet there is nothing it is like to be one of them; for them, there are no experiences; phenomenally, they are completely inert. These zombies, as Chalmers notes (p. 95), are different from Hollywood zombies who are psychologically and functionally impaired; our zombies are merely phenomenally impaired. My zombie, for example, is functionally identical to me, but is completely devoid of consciousness. This *does* seem logically coherent, unlike the cases for life, heat, and the rest. And this is the key. Zombies are logically possible. There is a possible world where the low-level physical facts are *exactly* like the facts in our world, where there is life, politics, and culture, but where there is no consciousness. Isn't it strange? MTV logically supervenes on the low-level physical facts of our world, but consciousness does not. In the zombie version of our world, they listen to MTV, but they don't experience any music. Their ears pick sound waves; their auditory systems process it, and their mouths say "O wow, The Fugees," but they don't feel the beat, they don't groove to the melody -- for them, it is all black on the inside.

Here now, in short form, is Chalmers' general argument against materialism (p. 123).

1. In our world, there are conscious experiences.
2. There is a logically possible world physically identical to ours in which the positive facts about consciousness in our world do not hold. [This is the *zombie world*].
3. Therefore, facts about consciousness are further facts about our world over and above the physical facts.
4. So materialism is false.

Accepting this argument turns on accepting the fact that consciousness does not logically supervene on the physical. To bolster the intuition that it doesn't supervene, here are Chalmers' five primary arguments for this conclusion (pp. 94-106).

Argument#1. The Logical Possibility of Zombies. Obviously, if Zombies are logically possible, then consciousness cannot logically supervene on the physical. But, obviously, zombies are logically possible, so.....

Argument#2. The inverted spectrum. If zombies aren't your cup of tea, then simply imagine someone physically identical to you but with *different* conscious experiences, e.g., your inverted twin sees a color spectrum inverted from yours. This again suffices to show that conscious experience doesn't

supervene on the physical. (This argument actually establishes that the specific character of conscious experience doesn't logically supervene, which is all Chalmers needs. It leaves open the possibility that the actual existence of consciousness does logically supervene. The zombie argument establishes the stronger conclusion.)

Argument#3. The surprise of consciousness. Consciousness is a surprising feature of the universe; we know about it only through our own experience. Even if we had a completed theory of cognition (and biochemistry, chemistry, and physics) that information would not lead us to postulate consciousness.

Argument#4. What Mary knew. This is closely related to Argument#3. Mary is the world's leading neuroscientist living in age of a completed neuroscience. But she grew up and now lives in a black-and-white room and lab. She knows all there is to know about the brain, but she has no idea what it is to see green (or any other color). The very first time she leaves her black-and-white room on a warm summer's day, she sees grass and experiences green for the first time. How could this be if color experience logically supervened on the physical? (See Jackson, 1982.)

Argument#5. The lack of any remotely plausible analysis of consciousness. Analyses in terms of, e.g., consciousness's functional role in the causal nexus of cognition won't work because consciousness's *functional role* isn't what's at issue; it is experience itself that is at issue.

All five of these arguments are intuition pumps. Taken together, they manage to pump quite a bit of intuition. Perhaps it is now plausible that consciousness doesn't logically supervene. Given that all other positive facts about the world *do* logically supervene, consciousness emerges as quite strange indeed, and in fact it requires us to reassess our metaphysical assumptions about the nature of the world. Beginning with chapter 4, Chalmers does just this.

Before we leave this section, I want to mention two objections that I hear a lot. Many people suggest that consciousness's not logically supervening is no cause for alarm nor the drastic measure of accepting dualism because *nothing* logically supervenes on the physical. The other objection I hear from some is that they genuinely cannot imagine a zombie world, nor an inverted spectrum world, nor the Mary situation. They agree that consciousness has yet to be explained, but when it is, we will see that it is reductively explainable, ineluctable, and not surprising at all, given the low-level facts. Chalmers' response to both of these groups is to urge them to be more careful about using their imaginations. Playing with possible worlds is no game

for amateurs. Presented with these two objections, he just carefully goes through the development of the intuitions trying to convince them of the error of their ways. I would suggest however, that the presence of these two *contrary* intuitions means that there is more to imagination and intuitions about what's possible than meets the eye. Both of these objections can't be true, but they can both be false. The fact that I have heard both advocated, as well as Chalmers' intuition that both objections are in fact false (he holds, to repeat, that one thing -- consciousness -- doesn't logically supervene), inclines me to think that intuitions about what is logically possible are very delicate, if not out right suspect. I will say more about this below. Still, in practice, I usually have found it somewhat easy to make plausible to the former objectors (if not convince them) that virtually everything *but* consciousness logically supervenes, and make plausible to the latter objectors that consciousness *alone* doesn't logically supervene. In any case, Chalmers' intuitions are not obviously false, and they lead to some interesting territory. Let's explore it.

#### 4. The Argument for Naturalistic Dualism

Naturalistic dualism is a property dualism. It is the thesis that consciousness is nomically related to physical events in the brain, but not reducible to them. Chalmers calls this nomic relation *natural supervenience* in contrast to logical supervenience, (see pp. 34-38). Natural supervenience uses the notion of natural possibility, rather than logical possibility. "A naturally possible situation is one that could actually occur in nature, without violating any natural laws" (p. 36). Chalmers holds that there is a science in the offing which will, if not reductively explain, at least systematically catalog the law-like relations between psychological states and phenomenal states. As Chalmers puts it: "To bring consciousness within the scope of a fundamental theory [beginning with physics and including everything] we need to introduce *new* fundamental properties and laws" (p. 126). Chalmers calls these laws whose existence he hypothesizes *psychophysical laws* (p. 127) -- though a better name would have been *phenophysical laws*, since for Chalmers psychological properties do logically supervene on the physical and therefore psychological laws are just like physical laws. Psychophysical laws are supervenience laws, specifying how conscious experience *naturally* (nominally) supervenes on the physical.

Naturalistic dualism (the topic of chapter 4) holds out the promise of a strange new addition to the world as we currently conceive it, but an addition that is still nomic and well-behaved. Since psychological states do logically supervene on the physical (on brain states, in fact), once the psychophysical laws are in place, we will have a more or less complete story from quarks to conscious experience and back again. It won't be a story involving reductive

explanation from top to bottom, but it will supply reductive explanations up to consciousness, and then psychophysical laws beyond that. Naturalistic dualism is depicted in figure 1.

Chalmers explains it nicely this way:

In a way, what is going on there with consciousness is analogous to what happened with electromagnetism in the nineteenth century. There had been an attempt to explain electromagnetic phenomena in terms of physical laws that were already understood, involving mechanical principles and the like, but this was unsuccessful. It turned out that to explain electromagnetic phenomena, features such as electromagnetic charge and electromagnetic forces had to be taken as fundamental, and Maxwell introduced new fundamental electromagnetic laws. Only this way could the phenomena [of electromagnetism] be explained. In the same way, to explain consciousness, the feature and laws of physical theory are not enough. For a theory of consciousness, new fundamental features and laws are needed.

This view [naturalistic dualism] is entirely compatible with a contemporary scientific worldview, and is entirely naturalistic. (p. 127)

Here is Chalmers' argument for naturalistic dualism (p. 161):

1. Conscious experience exists.
2. Conscious experience is not logically supervenient on the physical.
3. If there are [mental] phenomena that are not logically supervenient on the physical then materialism is false.
4. The physical domain is causally closed [this is a statement of the first law of thermodynamics].
5. Physically identical beings will have identical conscious experiences [I discuss this below when I discuss chapter 7].

Together, propositions 1, 2, and 3 entail proposition 6:

6. Materialism is false.

And propositions 4, 5, and 6 entail naturalistic dualism:

“...conscious experience arises from the physical according to some laws of nature, but is not itself physical.” (p. 161)

Chalmers does a superb job of cataloging each of the other major positions regarding consciousness (e.g., reductive functionalism, new-physics materialism, eliminativism, interactionist dualism), and showing how to get each position by denying one of the propositions 1 - 4. He then goes through each position, arguing that it suffers some irremediable flaw. What emerges then is that naturalistic dualism is the most tenable position to hold about consciousness. (see pp. 161 - 168).

## 5. Synopses of Remaining Chapters

So far, I have recounted the highlights up through chapter 4. In this section I will finish off the book, providing synopses of the rest of the chapters. In the next section, I will discuss an objection to Chalmers' view of the world.

Chapter 5 addresses the vexed problem Chalmers calls the *paradox of phenomenal judgment*. Here's the problem. If zombies are going to do the work Chalmers wants them to do, then they have to be psychologically identical to us (they are phenomenally as different as can be, however). But one of our *psychological* capacities is making judgments about our phenomenal states. How can zombies make judgments about phenomenal states if they don't have any? Yet they must make such judgments if they are to be psychologically identical to us.

Here's an example. Right now it is early morning and quite cold in my house. I am experiencing this coldness; I am conscious of being cold. So a nonphysical property of me is “associated with” a physical property. Because of my experience of cold, I come to believe I am cold, and I put on a sweater and vest and turn on the heat. This time, physical properties (my belief and subsequent plans and actions) are “associated with” phenomenal properties.

(We need a quick sidebar here. How this association is manifested is a deep mystery, and one that Chalmers considers in some detail, especially in chapter 8. It is the mystery that is supposed to be partially mitigated by psychophysical laws. Perhaps my phenomenal states are *caused* by my physical states. If so, then this would not be standard physical causation (which

we don't understand, anyway, see footnote 2). This causation would be just some sort of brute causation, and would still fall under the psychophysical laws. Chalmers hypothesizes that psychophysical laws have a special property: they do not violate the causal closure of the physical universe, i.e., they do not violate the first or second laws of thermodynamics. Therefore, they do not interfere with any laws in our physical universe (p. 127).)

To return to my example, note that my zombie twin behaves exactly like me. He gets up, shivers, grouses around looking for warm clothes, turns on the heat, and then sits down to write about it. My zombie twin even says: "It sure is cold in here, and it is making me feel cold." How can he do that when he isn't experiencing any cold? The situation is even worse than this. My zombie twin believes that his "experience" of cold is mysterious. He is right now writing a review of zombie Chalmers, and is therefore wondering how in the world *his* zombie twin (even though *he himself* is the real zombie) could form beliefs and judgments about experience when he (my zombie twin's twin) doesn't have any. My zombie twin thinks of his twin and says "Poor blighter, he has no conscious experience like I do, yet there he is diligently writing about it." My zombie twin must think these things because I do, and because such thoughts are purely psychological. But how can he do it? He lacks consciousness entirely; how can he think it is mysterious?

The situation is deeply weird, and Chalmers knows it. He wrestles mightily to subdue this paradox. But I have not met anyone who thinks he has succeeded. Still, it is good reading.

Chapter 6 begins Chalmers' development of a positive theory of consciousness: he begins fleshing out naturalistic dualism. This project turns on noting the tight, coherent relationship between cognition and consciousness, specifically the tight relationship between consciousness and awareness (awareness is a psychological property). The next step is unearthing principles governing this relationship. One such principle is the *principle of structural coherence* :

...the structure of consciousness is mirrored by the structure of awareness,  
and the structure of awareness is mirrored by the structure of consciousness.  
(p. 225)

Chalmers then discusses versions of this principle offered by other philosophers and this principle's relation to other theories of consciousness. Then he discusses the explanatory role of this principle, and finally the status of this principle as a universal law -- a psychophysical law.

Chapter 7 introduces the *principle of organizational invariance* :

... given any system that has conscious experience, then any [other] system that has the same fine-grained functional organization [as the first system] will have qualitatively identical experiences. (p. 249)

This principle entails that consciousness is an organizational invariant, i.e., that “functional organization fully determines conscious experience” (p. 274). This means that, in the actual world (not some possible world), any system functionally organized in a way identical to me will have the same sorts of conscious experience that I have. We will all see blue skies, orange fires, green grass, and red apples no matter what we are made of, as long as we are all functionally organized identically. (In chapter 7, Chalmers is dealing only with natural possibility or possibility in the actual world. This sort of possibility cannot violate the laws of nature in our world.)

This principle is unintuitive and not widely believed. Many philosophers think that absent qualia or inverted qualia are *empirically* possible and that this refutes the principle of organizational invariance. But it would be a *very* handy to have this principle if one wants to be a naturalistic, scientific, dualist. So Chalmers argues for it with a series of intuition pumps (aka, thought experiments). These intuition pumps establish, according to Chalmers, that absent qualia and inverted qualia are extremely implausible. Hence, the principle follows. Chalmers says:

[The arguments against absent and inverted qualia establish a weak form of functionalism] that I have called *nonreductive functionalism* , on which functional organization suffices for conscious experience with natural [i.e., not logical] necessity. On this view, conscious experience is determined by functional organization, but it need not be reducible to functional organization. (p. 275)

The principle of organizational invariance, like the principle of structural coherence, is a psychophysical law (but is probably is not a fundamental example of such a law, Chalmers says (p. 275).) Both principles are constraints on a future theory of consciousness built along the lines Chalmers envisions.

Chapter 8 is one of most fascinating chapters in the book. Here Chalmers develops a dual aspect theory of consciousness. The idea is this. The basic stuff of the universe is *information* . Information (in the actual world) has two aspects: a physical aspect and a phenomenal aspect.

The two aspects are, not surprisingly, related in law-like ways (these are the psychophysical laws).

There are several terrifically interesting things going in this chapter, I recommend it highly. For example, in this chapter we find an explanation for why zombies talk about consciousness (p. 291). Another eye-opener is the discussion concerning panpsychism. Since information is ubiquitous, then so must phenomenology. Even rocks contain systems that are conscious in some weak sense. All of this is calmly, methodically, and persuasively argued. This chapter is a good piece of philosophy.

Chalmers says that chapter 8 is very speculative, but it is a very necessary chapter if one is to have any hope of coming to understand and eventually accept his view of mind and the universe. It is in this chapter that one gets a view of how it all might come together: the irreducibility of consciousness, the respect for the laws of nature and science, the coherence between the two, and zombies talking of consciousness. That the theory developed in this chapter has some strange consequences is to be expected, actually. After all, it *is* consciousness Chalmers is trying to explain, and, though there is no agreement even on this issue, consciousness does seem to be one of the strangest facts about our world (the fact that there is no agreement on even whether it is strange seems to argue for its strangeness).

Chapter 9 is about conscious machines and strong artificial intelligence. Chalmers, perhaps surprisingly, and certainly refreshingly, is a computationalist and an advocate of AI. He discusses his version of computationalism and then combines it with his principle of organizational invariance to yield the result that thinking machines are not only possible, but they will be conscious, too. (This possibility, of course, is only natural possibility.)

I need to say a brief word about Chalmers' version of computationalism. His is much more metaphysical than mine and focuses on a notion of implementation, which he leaves unexplicated. I prefer a computationalism that is based on epistemology and explanatory goals (see Dietrich, 1989, 1990). In my version, implementations depend in large part on explanatory goals, so it is explanation that is central to my interpretation. This difference doesn't make a big difference here, and I think I can agree with many of Chalmers' conclusions in this chapter. Nevertheless, an explanatorily-based computationalism is preferable, I think, in part because it does justice to our practices in cognitive science.

Chapter 10 is about quantum mechanics and consciousness. Consciousness and quantum mechanics and been strange bedfellows for years (Penrose, 1989, is a good example of recent

speculations on their relationship). The situation here is reminiscent of the one between consciousness and causation. Just as consciousness and causation are perhaps usefully lumped together, perhaps consciousness and quantum mechanics should be also consolidated. Each is a deep puzzle on its own, but wrapped together, they might form one solvable problem. In fact, though Chalmers doesn't consider this idea in any detail, perhaps the three problems -- consciousness, causation, and quantum mechanics, should be wrapped together and solved as unit.

Here is the problem with quantum mechanics. On the one hand, quantum mechanics is a very successful mathematical description of the physics of elementary particles and processes. Quantum mechanics is arguably the best scientific theory we've got in the sense of its robustness in predicting outcomes of physics experiments. On the other hand, quantum mechanics is a completely unintuitive, virtually impossible to believe description of our world. In short, though the calculus of quantum mechanics is robust, the interpretation of that calculus is very problematic.

After a very clear and readable discussion of the options, Chalmers comes out for the Everett interpretation. I can't do the discussion justice here. Suffice it to say that, again, Chalmers has shown us how strange our world is, if only we take consciousness seriously.

## **6. An Objection**

I have an objection to Chalmers' theory of consciousness. This objection leapt out at me when I first read his book. After teaching a seminar centered around his book, I have a notebook full of objections, comments, and worries. I suspect many readers will have such notebooks, and I hope the best of the comments will get published. We can reasonably hope that the resulting conversation will allow us to make some progress on the problem of consciousness. Though it may not be the kind of progress Chalmers envisions, it will nevertheless be due in large part to his book.

My objection is this: What would it be like if consciousness *did* logically supervene? I think that careful analysis shows that the situation where consciousness logically supervenes would be exactly the same as the situation if Chalmers' arguments were correct. There is something question begging against the materialist when a "Yes" answer is given to the question "Would conscious experience be any different if it did logically supervene?" Chalmers sometimes talks as if taking consciousness seriously entails being a dualist. But can't materialists take consciousness

seriously, too? My point is that they can.

To begin, I want to note two things. First, the arguments against materialism from zombies and inverted spectra given above, and indeed all of Chalmers' "Doesn't Logically Supervene" arguments (see section 3, above), really amount to intuition pumps: they pump up what I will call our *Cartesian Intuition*. This is the intuition that our conscious experiences could be just what they are regardless of how the world is; that somehow our consciousness isn't really a part of the physical world. I think virtually everyone has this intuition. Chalmers exercises it. Chalmers uses this pumped-up Cartesian intuition to make the logical possibility of zombies plausible. Second, Chalmers' arguments for non-supervenience really rely on what *seems* possible. His argument is really this: "Zombies intuitively SEEM logically possible, i.e., zombies SEEM conceptually coherent, therefore they ARE logically possible; they ARE conceptually coherent."

This second point is crucial to the development of my objection. Chalmers (p. 66) believes that every conceivable world is logically possible (modulo some intricacies). For him, figuring out what is conceivable requires care. Nevertheless, what is conceivable begins with what *seems* conceivable. Chalmers agrees to this, in fact. He says that to make a conceivability judgment you consider a *conceivable situation* and then make sure you are describing it correctly (p. 67). But a conceivable situation is just something that seems conceivable. So if I can throw doubt on the move from the *seeming* conceivability of zombies to their *being* possible, then I will have undermined the Cartesian intuition in terms of what we are inclined to infer from it. When we are no longer confident in the Cartesian intuition, then Chalmers' main argument loses its crucial premise that there is a logically possible zombie world. Note that I am *not* trying to show that this premise is false. And, I am not trying to show that zombies do not seem logically possible; they clearly do. Rather I am trying to show that we can't be sure that the premise is true. We must be agnostics about its truth value. Given this, materialism is suggested merely because it is ontologically simpler.

Here is my strategy in developing my objection. The best move to make, I think, is simply to concede to Chalmers that we do in fact have the Cartesian intuition, but then show that conscious experience itself leads to this intuition. If this worked, then it would forcefully undermine our confidence in the Cartesian intuition, and hence our confidence in the intuition that zombies are possible, and hence, finally, our confidence in the argument that consciousness doesn't logically supervene. Note that we must show that consciousness itself leads to our Cartesian intuitions. I don't think it is in the cards to show that zombies are logically impossible. I

agree with Chalmers that they seem possible, and I don't think it is strong enough to just question in general the reliability of our intuitions about possible worlds (though I will do that, too). We must show that consciousness itself leads to our Cartesian intuitions.

Here's another way to put my strategy. Suppose it really was the case that consciousness logically supervened on the physical. So zombies are not in fact logically possible; neither are inverted spectra. Given this, would conscious experience be any different? If not, then perhaps any creature who was conscious, and smart enough, would have the Cartesian intuition and therefore *conclude* that consciousness doesn't logically supervene on the physical.

My argument will rely on an intuition pump of my own. At this level of philosophy, intuition is really all we have to go on. To get the reader's intuitions going my way, note that how things seem needn't be much of a guide as to how they are. We all know this, but it is worth stressing here. For Chalmers, conceivability is *the* guide to what is logically possible. But is this reliable? The history of mathematics is crawling with cases of mathematicians who thought something was logically possible which wasn't; indeed, most of the time what was thought logically possible was logically *impossible*. The attempt to prove Euclid's parallel postulate from the others postulates is a good case in point; the parallel postulate turned out to be logically independent of the other postulates (that is how we got nonEuclidean geometry). So just because something seems logically possible doesn't mean it is. Also, perhaps we should be dubious of the whole notion of a possible world. I have a beard. It seems plausible to me to say that there is no possible world where I do not have a beard because in the world with a beardless "Eric Dietrich," things would be so different that that gentleman wouldn't be me. Events clear back to the Big Bang would have to be different. Why suppose that *that* beardless fellow is me?

Chalmers, as already mentioned, insists that one make sure one is describing a possible world correctly when reasoning about possible worlds and making conceivability judgments. But taken together, the two points just discussed in the previous paragraph suggest that being careful isn't sufficient. What really matters when spinning tales of possible worlds is re-establishing epistemic contact with the actual world when you are done. This frequently requires quite a bit of patience (even if one is willing, as I am, to be liberal about the definition of the actual world), but it is the only check on conceivability that really works. Chalmers embraces this check, too. That is why he calls for a concerted search for psychophysical laws. My plaint here is that epistemic contact with the actual world is ambiguous between his view and the view that consciousness does logically supervene. Since that latter view is ontologically cleaner, we ought to embrace it.

Now for my core argument. Consider the case of technologically primitive rain forest dwellers. There is a small village of these people, and they have never had contact with technologically advanced cultures. Because of the density of the forest and their mode of life, the forest dwellers have never seen long parallel tracks or beams converging in the distance. One day they wake up to discover massive, twin steel pillars reaching up far into the sky, each pillar almost the size of their entire village. They look up into the blue sky and fluffy clouds and see the pillars coming together way up in the clouds. The question comes up of whether the pillars actually come together up there. After some debate it is agreed that, of course, the pillars come together, because, after all, that is what they see. Then one day, one preternaturally bright forest dweller after thinking hard about the eye and angles comes to the conclusion that the twin steel pillars might in fact be as far apart up in the sky as they are right here beside his village, and that the pillars only *seem* to come together b/c of the way the world works and the way the eye works. He takes his idea to the elders whereupon it is agreed that the village should remain agnostic about whether or not the pillars come together or not.

This tale shows exactly what I think is going on with consciousness. Consciousness might logically supervene on the physical, but we can't tell because of the nature of consciousness itself. Consciousness just *is* the experience of the physical world. What would you experience if you saw the process on which consciousness does logically supervene? You would just experience another quale. So you couldn't, by definition, see *that* quale cause or result in your conscious experience. It's *all* conscious experience.

Call the process on which our consciousness logically supervenes the "consciousness producing process," or CPP for short. CPP is part of what is called the *supervenience base* of consciousness (in general, if X supervenes on Y, Y is the supervenience base). What would you see ("see") if you saw your CPP result in your consciousness? You would experience just another quale: you would see some working neurons, for example. That is *not* seeing or experiencing the logical supervenience relation. It couldn't be. Why? Because experiencing CPP is not experiencing the supervenience base; it is not experiencing CPP *as* the supervenience base -- it couldn't be.

Let's approach my argument from another direction. Consider this question: How could we see ("see") CPP result in our consciousness? In order for us to see CPP actually result in our consciousness, we would have to see externally both the CPP and our own consciousness at the same time, *and* internally consciously experience our consciousness supervening. But it is

impossible to see our own consciousness externally and at the same time experience our consciousness supervening on something.

Here is another intuition pump for my conclusion using Searle's famous Chinese Room Argument (I'm just going to assume familiarity with that argument). It is the room's consciousness that logically supervenes on the person-in-the-room. But the room qua consciously experiencing entity cannot see its consciousness logically supervene on the person-in-the-room. The room could see the person, to be sure (suppose the room is made of plexiglass, and the room is looking in a mirror). But any phenomenal states the room has (which are due to the person-in-the-room; which supervene on what the person-in-the-room does), including phenomenal states *about* the person-in-the-room, are going to be just that -- phenomenal states. By definition, the room can't see those states result in its conscious experience because those states are not part of the supervenience base, they are, as I said, conscious experiences themselves. The room might be able to infer, in some sense, that what the person-in-the-room is doing results in its being conscious, but we can do that do, as I will now argue. That is, I shall now argue that we can plausibly infer that consciousness *does* logically supervene on the physical. This will complete my objection.

In general, to see X logically supervene Y, observer O has to be able to see both X and Y from an external perspective. O has to be able to see the relational properties of the stuff X is made of realized in the relational properties of the stuff Y is made of. But ultimately, the stuff X is made of (as well as Y) is characterized by its relation to us, and this is just more phenomenal experience, as we've seen. So everything in our external world is just some relation or other to something else. Question: what do all these relations relate? Answer: intrinsicness; i.e., there must be some intrinsic properties all these relations are relating. But consciousness, as Chalmers notes (p. 153), just is a case of an intrinsic, nonrelational property. Now it is reasonable to suppose that intrinsicness logically supervenes on extrinsicness (and vice versa?). Therefore, it seems reasonable that consciousness (intrinsicness) logically supervenes on the physical (extrinsicness). But, if my arguments above are correct, we can never directly experience this supervening relationship. And in fact, we can't even get any decent empirical evidence for it. We are lead to hypothesizing its existence by logical means alone.

In sum, here is what my objection comes down to. I agree with Chalmers that I can conceive of a zombie world. But even being very careful, I can also conceive of a world where consciousness logically supervenes and yet everything is exactly like it is here in this world -- specifically, in both worlds, the Cartesian intuition exists in all conscious, intelligent creatures. I

can conceive of Chalmers being right about zombies, and I can conceive of Chalmers being wrong about zombies because I can conceive of mistakenly inferring the existence of zombies merely by the existence of a logically supervening consciousness itself. At a minimum, this ought to make us very nervous about inferring what's possible from what we can conceive. And in the final analysis, if I had to pick which conceivable world to go with, I'd pick the world where consciousness does logically supervene, but by its very intrinsic nature, prevents us from seeing this fact.[ftnt3]

My objection amounts to a defense of materialism. This defense is expensive, because it entails that we cannot really know the truth about whether or not consciousness logically supervenes. In that sense, my objection, if it works, is very unsatisfying. However, though I can't go into details here, it may be that much of Chalmers' positive theory of consciousness can be used even if my objection is correct. This is because my objection partly uses his idea of the nature of intrinsicness and extrinsicness.

## 7. Conclusion

Chalmers has written a good philosophy book. I recommend it to everyone interested in the mind. There is a book out (of dubious merit, in my opinion) called *The End of Science* (by a writer at *Scientific American*, of all places (Horgan, 1996)). This tome argues that science is ending; we're coming to the close of scientific inquiry. Chalmers' *The Conscious Mind* is a great antidote for such claims. Even if Chalmers is only partly right, the world he has revealed is puzzling and exciting. It simply must be experienced.

## Endnotes

1. All references are to Chalmers unless otherwise indicated.
2. Strictly speaking, there are several things that do not logically supervene on the physical. Things like indexicality and negative facts do not, it seems (see pp. 81-89). These are not important to my discussion however, so I will ignore them. A more interesting case is the case of physical laws themselves. Physical laws and causation do not logically supervene on the physical facts. Why? Because there exists a possible world physically identical to our world but with

different physical laws. Imagine such laws as governing events of such rarity and obscurity that they never occur. So our worlds are indiscernible but they have different laws. The nonsupervenience of physical laws suggests that causation itself doesn't logically supervene: there is more to causation, it seems, than mere physical facts. (This is just to say that Hume's view of causation as constant regularity is unsatisfying somehow.) In short, we seem to be as confused about the nature of causation as we are about the nature of consciousness. And now an interesting idea presents itself: perhaps consciousness and causation are related in some deep way. Chalmers discusses this interesting idea on pp. 86, 152ff., and elsewhere. See also Rosenberg, 1996.

3. My objection is, if one squints, similar to an argument developed by Colin McGinn (1989). I highly recommend this paper.

## References

- Davidson, D. (1970). Mental Events. In L. Foster and J. Swanson, eds., *Experience and Theory*. London: Duckworth.
- Dietrich, E. (1989). Semantics and the computational paradigm in cognitive psychology. *Synthese*, 79 (1), 119-141.
- Dietrich, E. (1990). Computationalism. *Social Epistemology* 4 (2), 135-154. Reprinted in E. Dietrich (ed.) *Thinking computers and virtual persons*. San Diego, CA: Academic Press.
- Dietrich, E. (1995). AI and the mechanistic forces of darkness. *J. of Experi. and Theor. A.I.*
- Horgan, J. (1996). *The End of Science*.
- Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly* 32, 127-136.
- Kim, J. (1978). Supervenience and nomological incommensurables. *American Philosophical Quarterly* 15, 149-156.
- Kim, J. (1993). *Supervenience and Mind*. Cambridge: Cambridge Univ. Press.
- McGinn, C. (1989). Can we solve the mind-body problem? *Mind* XCIII (391), July.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review* 4, 435-50.
- Rosenberg, G. (1996). Consciousness and causation: Clues toward a double-aspect theory. Manuscript, Indiana Univ. Bloomington, IN.