The irreducibility of consciousness

Amy Kind
Claremont McKenna College

Abstract
In this paper, by analyzing the Chalmers-Searle debate about Chalmers' zombie thought experiment, I attempt to determine the implications that the irreducibility of consciousness has for the truth of materialism. While Chalmers claims that the irreducibility of consciousness forces us to embrace dualism, Searle claims that it has no deep metaphysical import and, in particular, that it is fully consistent with his materialist theory of mind. I argue that this disagreement hinges on the notion of physical identity in play in the discussion. Clarifying this notion in turn helps to clarify what it means to claim that consciousness is irreducible, and provides insight into the implications that the truth of this claim would have for the dualism-materialism debate. Ultimately, I suggest that the sort of irreducibility that can be motivated by the zombie thought experiment is not sufficient to require the rejection of materialism.

The question of whether there can be a physical reduction of consciousness once seemed to rest largely on the plausibility of two now-familiar thought experiments — Frank Jackson's Mary, the color scientist, and Thomas Nagel's bat. More recently, the focus of the irreducibility debate has shifted to David Chalmers' thought experiment about zombies, creatures who are physically indiscernible from humans but phenomenologically void. Confronted with this thought experiment, the materialist has appeared to have three options. She can reject the claim that zombies are conceivable, she can reject the claim that that the conceivability of zombies entails their possibility, or she can reject the claim that the possibility of zombies entails the irreducibility of consciousness. It has been generally assumed, however, that to accept all three of these claims — and thereby to accept the irreducibility of consciousness — would be to admit defeat.

In this paper, I want to examine the merits of this assumption that the irreducibility of consciousness is incompatible with materialism. I will come to this issue by way of John Searle's criticism of Chalmers' thought experiment. Both Chalmers and Searle are proponents of the irreducibility of consciousness, but, interestingly, they disagree about

its consequences.[1] While Chalmers claims that the irreducibility of consciousness forces us to embrace dualism, Searle claims that it has no deep metaphysical import and, in particular, that it is fully consistent with his materialist theory of mind.[2] Getting to the bottom of this philosophical disagreement will provide us not only with some important clarification about what it means to claim that consciousness is irreducible but also with some insight into the implications that the truth of this claim would have for the dualism-materialism debate. Ultimately, I will suggest that the sort of irreducibility that can be motivated by the zombie thought experiment is not sufficient to require the rejection of materialism.

I. Chalmers' naturalistic dualism vs. Searle's biological naturalism

The sense of reducibility in play in Chalmers' discussion of consciousness is that of *reductive explanation*. For Chalmers, a phenomenon is reductively explainable in physical terms if and only if it is logically supervenient on some lower-level physical properties.[3] A reductive explanation, though not necessarily fully illuminating, will nonetheless remove some of the mystery about the higher-level phenomenon in question by giving us an explanation of it wholly in terms of the simpler properties. The claim that consciousness is not reductively explainable thus depends on the claim that consciousness does not logically supervene on physical properties. Chalmers takes such properties to be the fundamental properties invoked by a completed

[1] In Chalmers (1996) and Searle (1992).

[2] Although Searle often seems to avoid the label 'materialist,' this is not because he thinks of himself as a dualist, but rather because he thinks that terms like 'dualist' and 'materialist' are pieces of an 'obsolete Cartesian vocabulary.' Searle (1997b, 162). According to Searle, materialism is often taken to imply the falsity of naïve mentalism. Though he thinks this implication is mistaken, he nonetheless refrains from calling himself a materialist in order to avoid association with antimentalism. See Searle (1992, 52-5).

[3] See Chalmers (1996, 48). He does qualify this claim, noting that it is not as clear that logical supervenience suffices for reductive explanation than that it is necessary for it. But rather than limiting himself only to the necessity claim, he prefers 'to note that there is *a* useful notion of reductive explanation such that logical supervenience is both necessary and sufficient.' (1996, 48)

theory of physics, with the added stipulation that a world's physical laws should be included in the supervenience base. I will return to this stipulation below, arguing that it is considerably more significant than Chalmers would have us believe; in particular, I will suggest that it turns out to be an important key to unraveling his disagreement with Searle. But it will be useful to postpone this discussion until we have before us a more complete picture of the disagreement.

The heart of Chalmers' case for the failure of logical supervenience lies in his argument concerning the logical possibility of zombies.[4] As I have noted, a philosophical zombie is a being physically indiscernible from a human being but phenomenologically void. My zombie twin, for example, has a brain and body that are physically identical to (i.e., physically indiscernible from) mine but has no phenomenally conscious states. There is nothing it is like to be my zombie twin. As Chalmers claims, and as many philosophers have agreed, there does not seem to be any logical contradiction in the notion of a philosophical zombie. In fact, there does not seem to be any logical contradiction in the notion of a whole zombie world, a world that is physically identical to the actual world though phenomenologically empty. But from the logical possibility of a zombie world it seems to follow directly that phenomenal properties do not logically supervene on physical properties, and the failure of logical supervenience in turn precludes any reductive explanation of phenomenal consciousness.

Given this argument, Chalmers views the rejection of materialism as only a short step away. The zombie world, in which there is no consciousness, is by hypothesis physically identical to the actual world. Thus, the fact that there is consciousness in the actual world must be a further fact, over and above all the physical facts. Materialism, which is committed to the claim that all facts are physical facts, is therefore false.

Having abandoned materialism, Chalmers adopts in its place a version of property dualism that he calls *naturalistic dualism*. Although he accepts the materialist claim that the only substances in the world are physical ones, he departs from materialism in claiming that some of those substances have phenomenal properties that cannot be reduc-

[4] Though he relies most heavily on the zombie argument, Chalmers gives four other arguments as well. For all five arguments, see Chalmers (1996, 93-106). Chalmers' zombie argument has been much discussed. For two interesting recent discussions, see Perry (2001) and Levine (2001).

tively explained in terms of physical properties. Importantly, however, the irreducibility of phenomenal properties entails only that they are logically independent of physical properties. We need not conclude that phenomenal properties are actually independent of physical properties, and in fact, Chalmers denies that they are. He devotes much of his book to developing a theory of the systematic dependence of phenomenal properties on physical properties in the actual world. In short, though he denies that consciousness supervenes logically on the physical, he nonetheless believes that it supervenes naturally on the physical (hence the *naturalistic* dualism).

To characterize Chalmers' view, then, we might say that in the actual world 'consciousness *arises* from a physical basis, even though it is not *entailed* by that basis.' (Chalmers 1996, 125) As Chalmers himself acknowledges, there is only a fine line between this purportedly dualistic view and the views of many of his opponents who claim to be materialists. Interestingly, in some cases there may not even seem to be any such line at all. In such cases, however, Chalmers claims that his opponents are fooling themselves about the ontological commitments of their theories.

One such opponent is Searle, who advocates a view that he calls *biological naturalism*:

> Consciousness, in short, is a biological feature of human and certain animal brains. It is caused by neurobiological processes and is as much a part of the natural biological order as any other biological features such as photosynthesis, digestion, or mitosis. (Searle 1992, 90)

According to Chalmers, Searle's view contains an implicit admission of dualism, since the claim that the brain causes consciousness suggests that consciousness is something over and above the brain. Thus, Chalmers claims that Searle should be categorized as a property dualist rather than as a materialist, 'despite Searle's own view of the matter.' (Chalmers 1996, 370, fn. 2)

Importantly, this *is* despite Searle's own view of the matter. Searle not only explicitly denies that his claims about the brain's causing consciousness commit him to property dualism, but he also steadfastly rejects the claim that the irreducibility of consciousness entails prop-

erty dualism.[5] Irreducibility, according to Searle, 'has no deep metaphysical consequences for the unity of our overall scientific world view.' (Searle 1992, 122)

This presents us with our puzzle. Chalmers and Searle both deny that consciousness is reducible to physical facts, and they both accept that consciousness arises from a physical basis. There thus seems to be nothing that could justify classifying Chalmers as a dualist but Searle as a materialist. Should we conclude that the dualist/materialist contrast is a mere terminological distinction without a substantive difference?[6] In fact, I think the answer to this question is 'no' — the disagreement between Chalmers and Searle is a nontrivial one, and it points to a fundamental unclarity in how we are to draw the dividing line between materialism and dualism.

II. Two senses of irreducibility

Perhaps the most natural response when presented with our puzzle would be to point a finger at the term 'irreducibility' as the source of the trouble. Indeed given how many different kinds of reduction abound in the philosophical literature it would not be surprising if it were to turn out that Chalmers and Searle have different things in mind when they claim that consciousness is irreducible. Searle himself separates five kinds of reduction: ontological, property-ontological, theoretical, logical, and causal.[7] What distinguishes these reductions from one another is the sort of reducing entity under consideration: objects, properties, theories, sentences, or causal powers. But all five of these types of reduction have something in common, namely, the idea of 'nothing but' — in general, when we reduce *A* to *B* we show that *A* is nothing but *B*, or that *A* consists in nothing more than *B* — and I think we can pretty quickly rule out the suggestion that the differences among these types of reductions matter for making sense of the Chalmers-Searle dispute.

[5] For his denial that he is a property dualist, see Searle (1992, 252, fn. 4); for his remarks about the irreducibility of consciousness, see e.g. Searle (1992, 116); Searle (1997c, 174).

[6] Chalmers considers, and rejects, the suggestion that their disagreement is merely terminological. See Chalmers (1996, 130).

[7] See Searle (1992, 112-116).

There is, however, another way in which an ambiguity in the notion of irreducibility might come into play. Chalmers, recall, is concerned primarily with reductive explanation, whereas Searle is interested in reduction. It will thus be useful to explore how Chalmers' notion of reductive explanation relates to the general notion of reduction.

First, notice that in order to reduce some phenomenon *A* to some other phenomenon *B*, we must be able to reductively explain *A* in terms of *B*. If we could not give a reductive explanation of *A* in terms of *B*, then facts about *A* could not logically supervene on facts about *B*. This would mean, however, that facts about *A* are over and above facts about *B*, and so we would not be able to reduce *A* to *B*. Thus, Chalmers' claim that consciousness cannot be reductively explained clearly entails the claim that there can be no reduction of consciousness in any of Searle's five senses.

The converse entailment, however, does not hold. While reductive explanation is required for reduction, reduction is not required for reductive explanation.[8] Consider the standard sort of case where reduction fails, namely, a phenomenon that is multiply realizable. If a phenomenon can be realized in many different physical substrates, then we cannot reduce it to any particular physical substrate. Temperature, for example, is realized differently in gases from how it is realized in plasmas; though the temperature of a gas is the mean kinetic energy of the molecules, plasmas do not even consist of molecules but rather of dissociated atoms. Nonetheless, the fact that temperature cannot be reduced to mean molecular kinetic energy does not prevent scientists from giving reductive explanations of the phenomenon.[9] What precludes the possibility of a reductive explanation of some phenomenon *A* in physical terms is the failure of *A* to supervene logically on the physical. But the mere fact that *A* is multiply realizable does not entail such a failure.

This suggests a solution to our puzzle. Although both Chalmers and Searle claim that

(a) Consciousness is irreducible

[8] Chalmers makes this latter point in Chalmers (1996, 43).

[9] For a discussion of the reducibility of temperature, see Churchland (1986, 356).

we now see that there are two nonequivalent ways of interpreting this claim:

(a1) There can be no reduction of consciousness
(a2) There can be no reductive explanation of consciousness.

Importantly, then, there are two different ways of interpreting Searle's claim that

(b) The irreducibility of consciousness has no antimaterialist consequences

namely,

(b1) There would be no antimaterialist consequences if consciousness were unable to be reduced.
(b2) There would be no antimaterialist consequences if consciousness were unable to be reductively explained.

Moreover, just as a1 does not imply a2, b1 does not imply b2. Since reduction is not required for reductive explanation, the claim that a phenomenon cannot be reductively explained is a stronger claim than the claim that the phenomenon cannot be reduced. There might be important metaphysical consequences were it the case that consciousness could not be reductively explained even if there would be no such consequences were it the case that there could be no reduction of consciousness.

We thus have an easy way to reconcile the seemingly incompatible claims of Chalmers and Searle. When Searle claims that the irreducibility of consciousness is consistent with materialism, he means only to be claiming b1, whereas when Chalmers denies that the irreducibility of consciousness is consistent with materialism, he means only to be denying b2. Though it seems that they disagree about the metaphysical implications of the irreducibility of consciousness, their disagreement really hinges on an ambiguity in the notion of irreducibility.

Unfortunately, however, we cannot take this easy way out. The proposed reconciliation suggests that Searle, like Chalmers, rejects b2. But if Searle admits that there would be anti-materialist consequences were it the case that consciousness could not be reductively explained, then in order to remain a materialist he would have to reject a2. Yet it is not clear that this avenue is open to him. The prob-

lem is that Searle seems to accept the conceivability of a zombie world; as he says, there is no contradiction in the assumption of 'a world where all the physical particles were exactly like ours, with a zombie doppelgänger for each of us, in which there was no consciousness at all.' (Searle 1997b, 147) Assuming that the conceivability of zombies precludes a reductive explanation of consciousness, as Chalmers has argued, Searle would be committed to a2.[10]

III. Zombies and living rocks

The above discussion rules out our suggestion that an ambiguity between reduction and reductive explanation was the source of the trouble. But importantly, it offers us an alternative suggestion, namely, that we might be able to trace the problem back to issues relating to zombies — either their conceivability or the implications thereof. Recall that the fact that we can conceive of a zombie world is supposed not only to preclude the possibility of a reductive explanation of consciousness, but also to entail the rejection of materialism. Chalmers' argument for the latter was straightforward:

(1) We can conceive of a zombie world (a world that is physically identical to our world but without any consciousness).
(2) Thus, a zombie world is logically possible.
(3) Thus, the fact that there is consciousness in our world is a further fact, over and above all the physical facts about our world.
(4) Thus, materialism is false.

Given that we have been interpreting Searle as accepting (1) but denying (4), we would be able to solve our puzzle were it the case that he rejects Chalmers' inference from (1) to (2), from (2) to (3), or from (3) to (4).

Searle's discussion of Chalmers' work draws our attention to the move from (2) to (3). It seems that he wants to deny that the logical possibility of a zombie world entitles us to the claim that, in the actual world, facts about consciousness are over and above the physical facts:

[10] One might think that there is another easy way out, namely, to treat Searle's use of the terms 'reduction' and 'irreducibility' as nonstandard (and indeed, some of his own remarks seem to support this suggestion). But the considerations I raise in the text will also count against this attempt to solve our puzzle.

> If I were to imagine a miraculous world in which the laws of nature are different, I can easily imagine a world which has the same microstructure as ours but has all sorts of different higher-level properties. I can imagine a world in which pigs can fly, and rocks are alive, for example. But the fact that I can imagine these science fiction cases does not show that life and acts of flying are not physical properties and events. (Searle 1997b, 147–8)[11]

Chalmers has responded by arguing that the passage just quoted contains a confusion:

> To show that flying is nonphysical, we would need to show that the world's physical structure is consistent with the *absence* of flying. From the fact that one can *add* flying pigs to the world, nothing follows. Second, the scenario he describes is not consistent. A world with flying pigs would have a lot of extra matter hovering meters above the earth, for example, so it could not possibly have the same physical structure as ours. Putting these points together: the idea of a world physically identical to ours but without flying, or without pigs, or without rocks, is self-contradictory. (Chalmers 1997, 164)

Chalmers should be granted this point. Nonetheless, it is not clear that his response adequately deals with Searle's suggestion. Even if we dismiss Searle's flying pig example as inconsistent, that example is only one of two that Searle gives. He also suggests that we can conceive of a world physically identical to our world in which rocks are alive — a suggestion that Chalmers' response ignores. Once we attend to this second example, we see that it cannot be as easily dismissed as the example of flying pigs. Though a world physically identical to ours but 'without flying, or without pigs, or without rocks' *is* self-contradictory, a world physically identical to ours but with *living* rocks is not — or, at least, not obviously so. And therein lies a problem for Chalmers. If we can conceive of a world physically identical to our world but which contains a rock endowed with life, then such a world is logically possible; by parity of reasoning with the zombie argument, we reach the unpalatable conclusion that the prop-

[11] Strictly speaking, this passage leaves upon the question of whether Searle is rejecting the inferences from (1) to (2) or from (2) to (3). However, Searle's own history of reliance on conceivability arguments strongly suggests that his objection must be to the latter of these two inferences.

erty of being alive fails to be reductively explainable in terms of physical facts.

In order to reject this claim about the irreducibility of life, Chalmers will have to argue that, contrary to how it might initially seem to us, we cannot really conceive of a world physically identical to ours in which rocks are alive. But given that his arguments for the irreducibility of consciousness and the failure of materialism depend on the conceivability of a zombie world, pursuing this line of response imposes on Chalmers an extremely delicate task. Any arguments to the effect that we are not really conceiving of what we think we are conceiving will have to apply to the case of the living-rock world without also applying to the case of the zombie world.

The difficulty of this task is exemplified by Chalmers' treatment of a similar objection to the zombie argument that arises from reflection on vitalism. Suppose the vitalists had offered the following argument to establish their position:

(1*) We can conceive of a lifeless world (a world that is physically identical to our world but without any living things).
(2*) Thus, a lifeless world is logically possible.
(3*) Thus, the fact that there is life in our world is a further fact, over and above all the physical facts about our world.
(4*) Thus, vitalism is true.

Vitalism, however, is surely false. Thus, unless we are willing to reject (1*), it looks as if there is something wrong with this argument structure.

Unsurprisingly, Chalmers wants to reject (1*). He attempts to justify this rejection primarily by reminding us of the original motivation for the vitalist view. The vitalist theory grew out of skepticism that we could explain all the complex functions associated with life (reproduction, behavioral adaptations, etc.) merely in terms of physical mechanisms. What the vitalists primarily sought, according to Chalmers, was an explanation of those functions, and so once such explanations were developed, 'vitalist doubts mostly melted away.' (Chalmers 1996, 109)

As a historical point, this is no doubt correct. When science developed an adequate theory of how physical processes account for the functions in question, vitalism died out. But all this shows is that people stopped believing that what accounted for life in the actual

world was some sort of *élan vital*. To convince us to reject (1*), Chalmers needs to show something stronger, namely, that there is something incoherent about the notion of a lifeless world. On this score, however, he has very little to say. The vitalism example in this way puts pressure on Chalmers' argument against materialism. Insofar as we do seem able to conceive of a lifeless world, despite the fact that we are committed to the physicality of life, there must be something wrong with the vitalism argument — and, correspondingly, with the zombie argument.

Notice, however, that the strength of the objection to Chalmers depends on the plausibility of the claim that a lifeless world is conceivable. We thus might reasonably think that Searle owes us some explanation of how exactly we are to conceive the lifeless world — or, to be fair, since Searle's own example was not of a world physically identical to ours with *no* life but rather of a world physically identical to ours with *extra* life, we might reasonably think that he owes us some further description of what exactly we are conceiving when we are conceiving the living-rock world.[12]

Once we look more closely at Searle's example, however, we find that his description of the living-rock world reveals what really separates the two philosophers. Importantly, when Searle indicates the respect in which that world is indiscernible from ours, he puts the point in terms of *microstructure*. The reason that he thinks that there is no logical contradiction in supposing that there could be a world with the same microstructure as ours in which rocks were alive is that he thinks microstructure alone is not sufficient for life. According to Searle, whether an object with a certain microstructure has the property of being alive depends on the laws of the world that the object inhabits. In our world, and in any world with our laws of nature, objects with the microstructure of rocks will not be alive. But he also thinks that there would be no logical contradiction in the supposition that, in a world that lacks our laws of nature, an object with a microstructure identical to an actual-world rock has the property of being

[12] In fact, this might be too charitable to Chalmers. When arguing for the logical possibility of zombies, Chalmers tries to shift the burden of proof to his opponents. There he claims that, in general, the burden of proof lies on the person who claims that a given description is logically *impossible*. (Chalmers 1996, 96) Thus, to make an *ad hominen* point, it might seem that the burden is on him to convince us that our supposition of a lifeless world contains a contradiction.

alive. Likewise, when Searle discusses Chalmers' zombie case, he again puts the point in terms of microstructure. He describes a zombie as 'physically identical to a normal human being down to the last molecule,' and, perhaps more tellingly, he describes the zombie world simply as one in which all the physical 'particles' are exactly like ours. (Searle 1997b, 147) Unfortunately, however, this crucially underdescribes the zombie world that Chalmers has asked us to imagine. For Chalmers, the zombie world does not just have the same physical microstructure — the same physical *particles* — as the actual world, but it also shares all of the physical laws of the actual world. As I noted in Section I above, Chalmers stipulates the inclusion of the physical laws of a world in the set of physical facts about it. By hypothesis, the zombie world is physically identical to the actual world, which means that the zombie world, by hypothesis, is one in which all the physical particles not only are the same as those in the actual world but also are governed by all the same physical laws as in the actual world.

IV. Two senses of physical identity

The preceding discussion has thus brought us to the heart of the matter, allowing us to pinpoint the issue between Chalmers and Searle. Though it looks as if they disagree about the implications of the conceivability of a zombie world, in fact they disagree over its conceivability. This disagreement is hidden, however, by an ambiguity in the notion of 'physically identical.' Let us explicitly distinguish the two different ways to understand 'physical identity':

> *Microphysical Identity:* Two worlds are microphysically identical if and only if they have the same physical microstructures.
>
> *Lawful Microphysical Identity:* Two worlds are lawfully microphysically identical if and only if they (a) have the same physical microstructures; and (b) have the same physical laws.

We can thus summarize the dispute as follows. While Searle requires only microphysical identity for two worlds to count as physically identical, Chalmers requires lawful microphysical identity for two worlds to count as physically identical.

Once we have drawn this distinction, the suspicion arises that Chalmers trades on this ambiguity in putting forth his zombie argu-

ment. Since the failure of local supervenience does not imply the failure of global supervenience, in order to show that consciousness cannot be reductively explained he needs to show (as he himself admits) that consciousness does not globally logically supervene on the physical facts. However, given that he thinks not only that it is easier, practically speaking, to defend the local version of the zombie argument than the global version, but also that 'if consciousness supervenes at all, it almost certainly supervenes locally,' he generally relies on the local version of the argument, noting that the argument could be transformed into a global version with 'straightforward alterations' if need be. (Chalmers 1996, 93) The problem is that the local version of the zombie argument cannot sensibly employ lawful microphysical identity. Although I defined both kinds of physical identity in terms of worlds, microphysical identity can be quite naturally stretched to cover physical identity between individuals as well: two individuals are microphysically identical if and only if they have the same physical microstructures. In contrast, lawful microphysical identity cannot be similarly stretched — or at least, it cannot be stretched in a way that allows it to be usefully invoked in arguments involving local supervenience. Laws are properties of worlds, not of individuals in the worlds. Perhaps the best we can do to apply lawful microphysical identity to individuals would be to say that two individuals are lawfully microphysically identical if and only if (a) they have the same physical microstructures; and (b) the worlds in which they are located have the same physical laws.

If this is right, then in order to conceive of my zombie twin as lawfully microphysically identical to me, I would have to build details about the world in which she exists into my conceiving of her. And it is not hard to see that this violates the spirit of running the zombie argument locally. Suppose, to use one of Chalmers' examples, that in order to determine whether value supervenes locally on the physical I conceive of an exact physical replica of the Mona Lisa. Since the replica clearly will not have the same value as the Mona Lisa itself, my thought experiment shows that value cannot supervene locally on physical properties. The value of the Mona Lisa depends on something besides its physical constitution, namely, its origin — it matters to the value of a painting who painted it. As Chalmers says, 'In general, local supervenience of a property on the physical fails if that property is somehow context-dependent — that is, if an object's possession of that property depends not only on the object's physical constitution

but also on its environment and its history.' (Chalmers 1996, 34) Notice what this suggests. When conceiving of a physically identical replica of some actual world object (or person) for the purpose of evaluating whether some property supervenes locally on the physical, we are required to conceive of the replica *independent of its context*. This requirement prevents my conceiving of a zombie replica that is lawfully microphysically identical to me, for to conceive of this, I have to build in to my conceiving facts about the laws that hold in the world where the replica exists, facts not about its constitution but (at least broadly speaking) about its environment.

When we are operating at the level of an individual zombie — which is the level at which Chalmers asks us to operate — the claim that a zombie is 'physically identical' to a conscious being must thus be a claim about microphysical identity. Importantly, as we have seen, this is a claim with which Searle agrees. He admits the conceivability of a being that is phenomenologically void despite being microphysically identical to a conscious being. But now, what happens when we move to the level of a zombie world, i.e., when we switch the argument from local to global supervenience? Doing so was supposed to involve only straightforward alterations but, if we are to conceive what Chalmers wants us to conceive, the alterations are not so straightforward after all. It is not just a matter of increasing the zombie population, that is, we are not just being asked to conceive of an entire world of these beings, each of whom is microphysically identical to a conscious being in the actual world. Rather, Chalmers wants the world itself to be lawfully microphysically identical to the actual world.

V. Irreducibility, revisited

Let us recall the puzzle with which we began. Despite their apparent agreement about the irreducibility of consciousness, Chalmers classified himself a dualist while Searle classified himself a materialist. We can now see that Chalmers and Searle do not, in fact, agree that consciousness is irreducible — or at least not in the same sense of irreducibility. The difference between microphysical and lawful microphysical identity infects their claims about irreducibility. Chalmers is committed to the claim that consciousness is what we might call *lawfully microphyscially irreducible*, i.e., that it cannot be reduced to physical properties plus physical laws. At best, however,

Searle is committed only to the claim that consciousness is what we might call *microphysically irreducible*, i.e., that it cannot be reduced to physical properties. And, since this difference seems enough to justify the fact that Chalmers and Searle fall on different sides of the dualist-materialist divide, we have solved our puzzle.

In doing so, however, I think we do more than simply settle an exegetical question about the theories of Chalmers and Searle. Rather, we are given some illumination about the significance of the debate about irreducibility and, perhaps even more importantly, about the nature of the divide between dualism and materialism. In this last section of this paper, I would like to discuss some of the lessons that I think we can learn from the preceding discussion.

The debate between dualism and materialism is often cast in terms of the issue of the physicality of mental states. To the question, 'Are mental states physical?,' dualists answer 'no' while materialists answer 'yes.' Often, the focus of the question narrows, and it is cast specifically in terms of consciousness: 'Is consciousness physical?' But again, an answer to this question is supposed to dictate on which side of the dualist-materialist divide one stands.

One tentative conclusion that can be drawn from our discussion of Chalmers and Searle is that this way of casting things is at best misleading (and at worst, simply a mistake). Once again, the problem lies in the distinction that we have previously uncovered between microphysical identity and lawful microphysical identity. The question, 'Is consciousness physical?,' also admits of two readings, depending on whether we understand physicality in terms of physical microstructure plus physical laws or simply in terms of physical microstructure alone. My own sense is that, though the form of the question itself encourages the latter interpretation, to use this question to draw the divide between dualism and materialism we would need the former interpretation. It does not seem to be enough for the dualist to deny that consciousness reduces to physical microstructure. If the reason that consciousness does not reduce to physical microstructure is that it reduces to physical microstructure plus physical laws, then it seems as if we are still, at least broadly speaking, within a materialist view. Dualism, in short, requires lawful microphysical irreducibility.

What implications does this have for the contours of a dualist view? At this point, it might be useful to look briefly at the view that Chalmers advocates. Having argued that consciousness does not logically supervene on the physical, Chalmers argues that it is none-

theless plausible to suppose that it *naturally* supervenes on the physical. In any world with our natural laws, a being that has the same physical microstructure as a conscious being will also be conscious. (Note that this means that the zombie world must not be one in which our natural laws hold. Moreover, since the zombie world has all the same physical laws as the actual world, this also implies that some of our natural laws are not physical laws.) In short, on Chalmers' theory, phenomenal facts have a strong, law-governed dependence on the physical facts. Nonetheless, he thinks that to account for these phenomenal facts we must postulate new (i.e., nonphysical) fundamental properties. Chalmers remains neutral on whether these new fundamental properties are phenomenal properties themselves, or protophenomenal properties on which the phenomenal properties supervene. In either case, however, he claims there will be new laws specifying the dependence of these new properties on the physical. These new laws — what he calls the *psychophysical* laws — fall outside the realm of physics.

But now the following question arises. Why do we need to posit both new properties *and* new laws in order to accommodate the lawful microphysical irreducibility of consciousness? Granted, once we posit the new properties, we might need to posit new laws that govern them, but the question I mean to be raising is whether we need to posit the new properties at all. Even if we accept Chalmers' arguments about the lawful microphysical irreducibility of consciousness, one might try to deny that we are forced this far. All such arguments show is that physical properties and the laws of physics are not enough to guarantee the existence of consciousness. But why, in order to account for consciousness, must we admit new, nonphysical properties? Why not simply admit new laws that are outside the realm of physics?

I suspect that Chalmers would have a ready answer to this question, namely, that unless there were new, nonphysical properties, nothing prevents the new laws from counting as laws of physics. But this answer leads us to an interesting result, one that I think puts pressure on Chalmers' argument for the naturalistic component of his dualism. His argument depends on the claim that we can conceive of a world in which all the physical particles are the same as those in the actual world, and in which all the laws of physics hold, and yet in which there is no consciousness. But, given his commitment to the natural supervenience of the phenomenal on the physical, he must

think that if we try to strengthen the thought experiment such that all the laws of *nature* hold in that world, we cannot conceive of that world as lacking consciousness. For example, if I try to conceive of my physical particles and structure being replicated by some creature in the actual world, I should be forced to see such a creature as conscious. The laws of nature that are not laws of physics thus play a major role in the conceivability of the relevant worlds or individuals. Notice, however, that there is no content given to the claim that something is a law of nature that is not a law of physics. All we can say about such a law is that it governs the (nonphysical) phenomenal properties. Importantly, we have no grasp of any such laws independent of the introduction of these phenomenal properties. And that, I would suggest, makes our intuitions about what we can and cannot conceive — at least with respect to the laws of nature — significantly less reliable.

It might appear we are on safer ground when it comes to thought experiments involving the laws of physics. But this appearance is deceptive. The laws of physics, by stipulation, consist not of the laws of current physics but of the laws of a completed physics. That means that when we conceive of the zombie world, we are supposed to be conceiving of a world in which all the laws of a completed physics hold. But in conceiving of the zombie world, we should not assume that these laws of physics exhaust the laws of nature; to do so would beg the question. Now, with all this on the table, I must confess that my confidence that I can conceive of what I am supposed to conceive has been significantly shaken. Is there a contradiction lurking in this supposition? Given the incompleteness of current physics, I do not see how anyone (even someone considerably more versed in the current state of physics than I am) could claim to be able to tell one way or the other.

So, where are we? Ultimately, once we draw the distinction between microphysical irreducibility and lawful microphysical irreducibility, then assuming (as I have been) that dualism requires the latter, the conceivability arguments in its favor are deprived of much of their intuitive force. Perhaps the dualist will retreat to the claim that his view requires only microphysical irreducibility. If this is the case, however, then the contrast between dualism and materialism is, I

think, deprived of much of its interest. If consciousness is merely microphysically irreducible, then we can draw the conclusion that the existence of consciousness depends very closely on the laws of physics. And, importantly, that seems something with which the materialist could well be content.

Amy Kind
Claremont McKenna College
Department of Philosophy
Claremont, California, United States
amy.kind@claremontmckenna.edu

References

Chalmers, David J. 1997. 'Response to 'Consciousness and the Philosophers'.' *NewYork Review of Books*. Reprinted in Searle 1997a.

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

Churchland, Patricia. 1986. *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, Mass.: The MIT Press.

Levine, Joseph. 2001. *Purple Haze.* Oxford: Oxford University Press.

Perry, John. 2001. *Knowledge, Possibility, and Consciousness.* Cambridge, Mass.: The MIT Press.

Searle, John. 1997a. *The Mystery of Consciousness*. New York: New York Review of Books.

Searle, John. 1997b. 'Consciousness and the Philosophers.' *NewYork Review of Books*. Reprinted in Searle 1997a.

Searle, John. 1997c. 'Response to David Chalmers.' *NewYork Review of Books*. Reprinted in Searle 1997a.

Searle, John. 1992. *The Rediscovery of the Mind*. Cambridge, Mass.: The MIT Press.