Dietrich, E. and A. Gillies (2001). Consciousness and the limits of our imaginations. *Synthese* v. 126, n. 3, pp. 361-381.

# Consciousness and the Limits of Our Imaginations

Eric Dietrich
Philosophy Dept.
Binghamton University
Binghamton, NY 13902

Anthony Gillies
(then of Dept. of Philosophy
Univ. of Arizona
Tucson, AZ, 85745). Now at
Univ. Michigan

**Consciousness and the Limits of Our Imaginations**

I. INTRODUCTION.

Reductive materialism about the conscious mind seems a reasonable theory, given the success of our other reductive materialist explanations: life is biochemistry, heat is mean molecular motion, light is electromagnetism, phenotypic traits are DNA sequences, evolutionary change is natural selection, and so forth. For the purposes of explanation, the entire universe of macrophysical facts seems to be cleanly reducible to lower-lever microphysical facts. This suggests that consciousness, too, eventually will be reduced. Unfortunately, at the current stage of cognitive science, this sort of inductive maneuver is the best argument we have for materialism; put bluntly, we do not know how it is that a brain could be conscious.[1] This, of course, makes materialism about the conscious mind completely different from the other "materialisms". Arguments for them stand alone, and are examples of our best reductive scientific explanations. On the other hand, not only do we have no reductive explanation in the case of consciousness, it is not even clear what one would look like. The cold truth is that materialism survives today on a wing and prayer.

And it may not survive too much longer if David Chalmers' arguments carry the day. In *The Conscious Mind*, Chalmers argues that reductive materialism is false.[2] In its stead he wants to put what he calls *naturalistic dualism*. His anti-materialist arguments are an interesting twist on a well-known argument form, and his naturalistic dualism is exciting to contemplate. Nevertheless, we think we can save materialism from the Chalmerian attack. This is what we do in the present paper.

Here is a summary. Chalmers' anti-materialist argument is that consciousness does not logically supervene on the physical because we can coherently imagine a world physically identical to ours in which the inhabitants are not conscious (this is called the *zombie world*, and its inhabitants are our *zombie twins*). From this he concludes that there is more to conscious phenomena than mere physical facts, and therefore that materialism is false. Note the three critical maneuvers in this argument: first, it is modal-metaphysical; second, it assumes a tight connection between what is conceivable and what is possible[3]; and third, in order to make a case for the last move, Chalmers needs a solid account of conceptual truth in virtue of meanings.

---

[1] Of course, materialism seems to get a leg up on dualism because we do not know how dualism could be true. But as we just said, we do not know how materialism could be true either. So it seems that materialism and dualism are on a par, at least as far as explaining how they could be true is concerned. We intend to defend materialism, but we won't be explaining how it works.

[2] Chalmers, *The Conscious Mind* (New York: Oxford, 1996)

[3] See pp. 65-69 (esp. 66), and 73. We might also call this Chalmers' confidence in the *a priori*.

Our response to this argument is that we cannot trust our modal intuitions about the supervenience of consciousness. First, using Chalmers' own tools, we argue that we don't have zombie twins; zombie twins are not possible. So materialism is true; at least it is the inference to the best explanation[4]. We concede, however, that zombies seem possible, that is, we can conceive of them. In fact, the thought that zombies are possible is almost irresistible. So if our argument is right, then the conceivability of zombies isn't sufficient for their possibility, and there isn't a tight connection between what is conceivable and what is possible as far as consciousness is concerned. Hence, materialism is true, even though it doesn't seem true.

At this point, our argument cries out for an explanation of why zombies *seem* possible. Their conceivability is the strongest objection to our anti-zombie argument, and, of course, one of the strongest objections to materialism. So, if we could explain why zombies are conceivable, even if they aren't possible, we would remove this objection, and hence, we wouldn't have to take zombie conceivability into consideration when it comes to evaluating the truth of materialism.

We offer such an explanation using consciousness itself. We show that the very nature of consciousness itself precludes us from ever directly determining whether consciousness logically supervenes or not. In other words, because we are conscious, even if consciousness did logically supervene, the conceivability of zombies is virtually irresistible. (We use Searle's Chinese Room to help us make our case.) This means that conceivability isn't a guide to what's possible, and that when it comes to consciousness, the "from-the-armchair" approach advocated by Chalmers won't work.[5]

For the first part of our defense of materialism, we argue that Chalmers is best seen as a Kripkean: a possible world inhabited by a non-conscious David Chalmers (a possible world with Chalmers' zombie twin) is a counterfactual situation about David Chalmers.[6] We then argue that we cannot coherently imagine ourselves without consciousness because when we affirm our own consciousness (taking something like "I am conscious" as paradigmatic), this, if true at all, is a conceptual truth on the very picture that Chalmers gives. We conclude, starting from Chalmers' own foundations,

---

[4] We see ourselves as reinstating the default position on mind-body problem, namely, materialism. But technically, the impossibility of zombie twins shows that either materialism is true or a kind of dualism (nonmaterialism) is true where the physical and the nonphysical are logically tied together. This type of dualism is not Chalmers' type at all. We regard this latter kind of dualism as much less plausible than materialism, so with Chalmers' zombies out of the way, we can embrace materialism with reasonable confidence.

[5] This underscores the point that "[c]onceivability is no proof of possibility" made by Putnam, "Meaning and Reference," *Journal of Philosophy* 70 (1973): 699-711.

[6] Kripke, *Naming and Necessity* (Cambridge: Harvard, 1972, 1980).

that zombie twins are not possible.

But why do zombies seem conceivable, at least to many? In the second part (section 5), we use a modified Systems Reply to Searle's Chinese Room Argument[7] to argue that we can never be in a position to see directly the mechanisms which cause our conscious experience, even if consciousness were to logically supervene on those mechanisms (the same point works for weaker forms of supervenience). So we cannot conclude from the fact that consciousness seems to have an ontologically distinct and physically irreducible status that it *is* ontologically distinct.[8] With our zombie intuitions explained away, the way is cleared for our anti-zombie argument (and other arguments, like the inductive argument with which we began), and so we conclude that materialism is true, even though it doesn't seem to be, and even though we can't provide a reductive explanation for it, at least in the foreseeable future.

Readers might well wonder why we can't stop with showing that it is impossible for us to have zombie twins. Isn't that enough to defend materialism? Yes, assuming as we did in footnote 4 that a dualism with logical glue isn't plausible. But what really needs to be done is to *explain why* it seems that zombies are possible -- *why* we can conceive of them. This leads into the realm of intuitions and what one imagines is logically possible.

The debate about phenomenal consciousness has never really been about whether *in fact* zombies are possible or not, i.e., whether *in fact* consciousness is logically independent of the physical. For the same evidence is available to all: we each know that we are conscious and aren't zombies, and we each believe for complex inductive reasons that others are conscious, too; yet there is wide disagreement about the possibility of zombies. No, the debate is not about the facts, rather, the debate has always been about intuitions: what we can *imagine*; what we *think* is logically possible. (Or to phrase it for the realists, the debate about whether in fact there are zombies always reduces to what we can imagine or conceive.) Presented with our argument that zombie twins are not possible, Chalmers and others of his ilk will likely simply reply that we have not been careful enough in describing the situation, e.g., we might have misdescribed or misidentified the relevant intensions[9]. So, the debate is whether we have been careful enough in our description, and this is a debate about intuitions, not about the facts. If we can explain away our intuitions that zombies are possible in a way that is completely compatible with materialism, then we will have completely

---

[7] Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences* **3** (1980): 417-424.

[8] Our argument is similar to one developed by Colin McGinn (*Mind* XCVIII, 391 (July 1989): 349-366).

[9] Chalmers, pp. 131ff.

undermined our faith in our modal intuitions in so far as consciousness is concerned.

## II. CHALMERS' ANTIREDUCTIONIST ARGUMENT

By 'consciousness' we will here mean exactly what Chalmers means: our phenomenological experience of the world. 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 (qualia); and it is experiences, qualia, that give us our subjective point of view. Being conscious is what makes it fun or horrible or merely boring to be a human.

Conscious experience is the most familiar thing in the world. You know nothing as certainly as your own conscious experiences (which are 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 little illuminating about consciousness. This may be one of the many spots where philosophers and psychologists will disagree with Chalmers, but the statement is true nevertheless. If science had managed to say something substantial about consciousness, not just a few individual scientists here and there, 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. And according to Chalmers, there is a very good reason for this: every science we have probes some aspect of the physical universe, and the knowledge our sciences provide us is physically mediated. But consciousness is not a part of the physical universe. Our knowledge of it is direct and unmediated. So in an important sense, though it is walled off from scientific inquiry in a traditional sense, it is not from our individual epistemologies.[10]

In arguing against materialism and for dualism, Chalmers does not use the notion of identity directly, e.g., psychophysical identity is never used at all. Instead, Chalmers relies on supervenience.[11] The key notion of supervenience is *logical*

---

[10] Chalmers says "...conscious experience is at the very center of our epistemic universe." Ibid., p. 74.

[11] Davidson, "Mental Events," in Foster and Swanson (eds.), *Experience and Theory* (London:

*supervenience*. This is the only kind of supervenience we will be concerned with here. The definition of logical supervenience is:

> 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.[12]

Everything in the physical world logically supervenes on the level below it (down to some basic, fundamental level). Fix these 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 atomic facts, the chemical facts, the biological facts, the psychological facts, and the social and cultural facts. In other words, it is logically impossible that there be 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. And therefore, you cannot coherently imagine such a world.

Here's a simple 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. If you think you can, you are mistaken.[13] For, according to Chalmers, 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.

Now extrapolate from this example to the level of our entire universe. It is simply impossible 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. In short, and though it may sound strange, MTV logically supervenes on the microphysical facts of our world. There is no possible world with the same microphysical make-up 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.[14]

---

Duckworth, 1970); and Kim, *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993).

[12] *The Conscious Mind*, p.33.

[13] *Ibid*, p.109.

[14] 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 *ibid*, pp. 81-89). These are not important to our

What is the Chalmerian argument that consciousness doesn't logically supervene? The argument turns on the fact that we can imagine zombie twins. 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, are different from Hollywood zombies who are psychologically and functionally impaired; the zombies which cause trouble for materialism are merely phenomenally impaired.[15] Chalmers' zombie twin, for example, is functionally identical to Chalmers, 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. Zombie twins are logically possible (or at least conceivable -- we return to this crucial point below). Hence, 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 -- the life, politics, and culture is the life, politics, and culture of zombies. It seems odd that MTV logically supervenes on the low-level physical facts of our world, but consciousness does not; odd, but, according to Chalmers, true nevertheless. In the zombie version of our world, they listen to MTV, but they don't experience any music. Their ears pick up 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.

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.

---

discussion however, so we 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 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 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, "Consciousness and causation: clues toward a double-aspect theory", Manuscript, Indiana University 1996.

[15] *The Conscious Mind*, p.95.

4. So materialism is false.[16]

Chalmers gives a total of five arguments that consciousness doesn't logically supervene. For completeness, we list them here.

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?[17]

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.[18]

All five of these arguments are intuition pumps. Taken together, they manage

---

[16] *Ibid,* p.123

[17] Jackson, F. "Epiphenomenal Qualia", *Phil. Quarterly* **32** (1982) 127-136.

[18] *Ibid.* p. 94-106

to pump quite a bit of intuition. Perhaps it is now plausible that consciousness doesn't logically supervene. Given that all the 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. Chalmers' reassessment results in his theory of naturalistic dualism. (We will not discuss this here, but refer the reader to chapter 4 of his book.)

For the purposes of this paper, we will only address his zombie twin thought experiment. It is both the most compelling and the most central of his arguments. The two crucial parts of this argument which will occupy us are (1) his use of possible worlds, and (2) his general semantic framework (in sections III. and IV., respectively).

## III. THE KRIPKEAN VIEW OF WORLDS

Transworld identity is a crucial issue in modal metaphysics: what is the criterion for identifying individuals across possible worlds? It is our view that the answer one gives to the problem of transworld identification is a function of how one settles the general ontological question about possible worlds. For instance, Hintikka does not appear to be bothered by the problem at all, because he takes 'possible worlds' terminology metaphorically. Yet, Hintikka has realized to his "considerable consternation that the likes of David Lewis and Alvin Plantinga [both of whom do see a problem to be solved] are taking the metaphor literally."[19]

Let's take a closer look at how these two issues, the ontology of possible worlds and the transworld identity problem, are related. In general, we can take our modal commitments one of two ways: (1) realistically, in which case we take our modal talk to refer to existent possibilia; or (2) nominally, in which case our modal talk does not refer to the existence of ways things might have been but rather represents turns of phrase (i.e., a metaphor to help cash out a particular formal apparatus). Hintikka appears to be in the latter group. Chalmers, however, is some sort of modal realist. In particular, he says that our formalisms for capturing modalities are logically dependent upon our modal-metaphysical intuitions.[20]

There are, of course, varieties of modal realism. If we exclude Kripke's view for the moment, realists minimally contend that sentences are necessary or contingent depending on whether the sentence is true at all possible worlds or true at some world

---

[19] Hintikka, "On Proper (Popper?) and Improper Uses of Information in Epistemology," *Theoria* LIX (1993): 158-165.

[20] *The Conscious Mind*, pp. 35, 362n6; and personal communication.

or other, respectively. This, on the realist view, in turn commits us to admitting independent entities (viz., worlds) into our ontology. It is the independent existence of such entities that allows our modal talk to be genuinely referential. Lewis' variety of realism, dubbed "Lewisian mad-dog realism" by John O'Leary-Hawthorne[21] , holds that possible worlds are concrete physical counter-parts to our own but that they exist in an alternative space-time dimension.[22] This is clearly the extreme case, but the point to be made holds for the entire class of modal realists. As soon as we admit worlds as independent entities, then we must face the epistemological issues of that conviction: how do we know what these possible worlds are like? In particular, how can we tell who's who in a possible world? Put more perspicuously, once we commit ourselves to modal realism, the problem of transworld identity arises. And now the proponent of modal realism owes us some solution or other.

Chalmers does not want to talk about identification across possible worlds. Thus he writes that, in general, he

> will not be concerned with questions about whether individuals in those [other possible] worlds might have different "identities"….These issues of transworld identity raise many interesting issues, but are largely irrelevant to my uses of the possible-worlds framework.[23]

Since Chalmers does not think he owes us a solution, it follows that he does not think the problem of transworld identity is relevant to his enterprise. If he does not find this problem relevant to his project, then (since modal realism is what makes the problem meaningful) we have that Chalmers is not a typical modal realist about possible worlds. This is more than mere infelicity: we know that Chalmers wants to be a realist of some sort with respect to our modal talk, since our modal intuitions are basic for him; but on the other hand, given his reluctance to address transworld identity, we have that he is not a realist about possible worlds.

Now Chalmers needs a way out. Is there a sort of realism that avoids the problem of transworld identity? Kripke, it seems to us, is just this sort of realist.

That Kripke is some sort of realist when it comes to understanding our modal discourse is easily verified. In explaining how it is he understands "worlds talk", Kripke writes:

---

[21] "The Epistemology of Possible Worlds," *Philosophical Studies* **84** (1996): 183-202. The reference to Lewis' flavor of realism is on p.198.

[22] Lewis, *On the Plurality of Worlds* (Oxford: Blackwell, 1986).

[23] *The Conscious Mind*, p.367n30.

> But I do not wish to leave any exaggerated impression that I repudiate possible worlds altogether, or even that I regard them as a mere formal device. My own use of them should have been extensive enough to preclude any misunderstandings.[24]

Given that possible worlds are more than mere formalism for Kripke, it is clear that he is not a modal nominalist. Kripke's brand of realism, to be sure, is the most modest strain. He makes the distinction between his interpretation and other modal realists in terms of an analogy to elementary probability. Given two distinguishable dice A and B, each fair and six-sided, the probability of throwing an eleven is easily computed. Eleven could turn up in exactly two ways: die A is 6 and B is 5, or A is 5 and die B is 6. Since each die has six sides, $P(11) = 2/36 = 1/18$. In order to figure this and similar probabilities, we discuss possible outcomes quite routinely. Kripke's point is that we can do this, and do it meaningfully, without positing "that there are some thirty-five other entities, existent in some never-never land, corresponding to the physical object before me."[25]

For Kripkean realism of the sort we have been discussing there is no problem of transworld identity to solve. If we are considering some counterfactual situation about, say, John Nash—perhaps one in which he never studied mathematics and never won the Nobel Prize—there is no transworld identity problem. Witness: there is no issue of us traveling to a distant galaxy in which someone resembling John Nash in certain ways never studied mathematics, and forcing us to ask "How do we know that *this* is John Nash?" Rather, by virtue of the fact that the counterfactual situation under consideration is one about *John Nash*, the problem just dissipates. This is precisely what Kripke means when he says that "'[p]ossible worlds' are stipulated, not discovered by powerful telescopes."[26] So, following the Kripkean line, there is no problem in stipulating that a discussion regarding how things might have been for Nash in some contrary to fact scenario is still a discussion about Nash in some contrary to fact scenario.

The Kripkean view of possibilia seems to fit Chalmers rather well. This is a way for Chalmers to embrace realism but eschew transworld identification altogether. The match is not quite as felicitous as it seems at first blush, however. The difficulty lies in the fact that Chalmers wants to describe worlds purely qualitatively, and Kripke

---

[24] *Naming and Necessity*, p.16.

[25] *Ibid.*, p.17.

[26] *Ibid.*, p.44. This defense of stipulation is often misinterpreted. A good discussion of it can be found in Nathan Salmon, "Trans-World Identification and Stipulation," *Philosophical Studies* **84** (1996): 203-223.

thinks this is an unnecessary restriction. Recall Chalmers' definition for logical supervenience. For the argument against materialism using the logical possibility of zombie twins, he wants to qualitatively specify a world identical to our own with respect to all microphysical facts, but differing with respect to phenomenal facts. In the characterization of supervenience that Chalmers gives, it is not clear that there are any rigid designators at all, and rigid designation is what is doing the work on the Kripkean view. This seems like an insuperable difficulty between the two positions, but in fact it is not.

For supervenience in general, it may well be the case that qualitative descriptions and their identities are what is in order. But, we think, the case for consciousness as Chalmers sees it is different. Remember that Chalmers takes conscious experience to be "at the very center of our epistemic universe."[27] It is from this center that his zombie thought-experiment originates. Since each of us can only know (directly, nearly incorrigibly, etc.) about our own private phenomenal experiences, Chalmers cannot ask us to imagine a world physically identical to ours (in the qualitative sense, of course) but in which everyone else lacks consciousness. Modulo our own private experiences, that could be this world for all we know. Rather, he is asking each of us to imagine our own zombie twin: someone qualitatively identical to ourselves, but who lacks phenomenal experience. So, the anti-reductionist argument is not "imagine all the same A-facts, without the B-facts holding", as Chalmers mistakenly poses it, but instead "imagine a possible world inhabited by your zombie twin, identical to you with respect to physical facts but a zombie nonetheless." We can easily imagine such a world inhabited only by our "twins" with varying degrees of similarity to ourselves, with one being physically identical without having phenomenal experiences. Suppose your zombie twin had a world-contemporaneous *non*-zombie twin. Which "twin" would Chalmers consider as counting for his thought experiment? Chalmers is likely to reply that, by stipulation, the one without phenomenal states is the one that matters. And this is precisely the Kripkean line.

So, when we are imagining zombie twins, we are simply imagining counterfactual situations about ourselves in which we do not have qualitative phenomenal experiences. And, taking names and demonstratives to designate rigidly, we then may conclude that the Kripkean view of stipulation is indeed correct for analyzing the Chalmerian supervenience relation in the case of consciousness.

---

[27] *Ibid.*, p.74.

So now we have the most charitable read of Chalmers' position as relying on Kripke's idea of stipulation, and thus at bottom his argument for the nonsupervenience of consciousness does not in fact rely on qualitative description of worlds.

To sum up, the aim of this section has been to establish the plausibility of taking the Chalmerian view of worlds to be largely Kripkean: possible worlds are counterfactual situations about the subject under discussion. When it comes to consciousness and zombie worlds, what Chalmers is really asking us to imagine is a counterfactual situation about ourselves in which we do not have phenomenal experience. Chalmers can resist this interpretation of his position only if he provides us with an alternative, and that would require providing us a solution to the transworld identity problem.

## IV. CONSCIOUSNESS AND CONCEPTUAL TRUTH

The step from the conceivability of zombie twins to their logical possibility is crucial for Chalmers. If he cannot tie these two together, then his thought experiments (e.g., imagining a zombie twin world) are of no consequence to materialist theories of mind. After outlining the case he makes for tying conceivability to possibility, we will argue that given the Kripkean account of worlds, the assimilation fails.

The traditional theory of meaning and reference, handed down since Frege, holds that for each concept there is some function $f : W \dashrightarrow R$, from the set of all worlds to the set of all referents, such that when fulfilled by the appropriate argument $w < W$, $f(w)$ yields an *extension* of the concept. The function itself is known as the *intension* of the concept. Chalmers wants to use Kripke's insight here:

> no single intension can do all the work that a meaning needs to do. The picture developed by Kripke complicates things [for the traditional view] by noting that reference in the actual world and in counterfactual possible worlds is determined by quite different mechanisms.[28]

From here, Chalmers splits the class of intensions into disjoint subclasses: primary intensions and secondary intensions, which correspond to the two ways in which reference is fixed. In the first case, we have reference as it is fixed in the actual world (primary intension), and in the second we have how reference is fixed in

[28] *Ibid.*, pp.56-7.

counterfactual situations given reference in the actual world (secondary intension).[29]

To illustrate the distinction, Chalmers gives an example.[30] Consider the concept "water." Suppose that the clear, drinkable liquid in the lakes, oceans, and rivers in the actual world turned out not to be H2O but something else entirely—perhaps Putnam's XYZ.[31] If the world had turned out that way, then "water" would pick out XYZ; but in our world it didn't, and thus "water" picks out H2O. Chalmers underscores this point by reiterating that "the primary intension of 'water' maps the XYZ world to XYZ, and the H2O world to H2O."[32] On the other hand, given that the clear drinkable liquid in the lakes, oceans, and rivers is H2O in the actual world, we can take the Kripkean line here and say that "water" picks out H2O in all possible worlds. Here, it is the secondary intension of "water" doing the referential work.

Whereas secondary intensions depend heavily on empirical factors, primary intensions do not. In the case of "water", the primary intension is something like "the clear, drinkable liquid in the lakes, oceans, and rivers", or as Chalmers puts it "watery stuff".[33] Already we can see that secondary intensions are intimately related to Kripke's necessary a posteriori truths, and that primary intensions cohere in the same way to a priori necessity. Indeed, "Kripkean a posteriori necessity arises just when the secondary intensions in a statement back a necessary proposition, but the primary intensions do not."[34] So, since "water" could have picked out XYZ in the actual world, "water is H2O" is not necessary using the primary intension of "water", but given that the actual world turned out the way it did "water is H2O" is necessary using the secondary intension of "water". Likewise, "water is watery stuff" is necessary when evaluated from the primary intension.

Earlier we made mention of Chalmers' faith in the a priori (in footnotes 3 and 29, above) . His surety in a prioricity stems largely from this distinction in intensions. And since his argument against reductive/functional explanations of the phenomenal

---

[29] The use of this two-dimensional semantic framework is problematic. In fact, we are quite skeptical about Chalmers' faith in the *a priori* as it is manifested through his use of primary intensions. A good critique of Chalmers' use of this framework (as well as Frank Jackson's) can be found in a recent (online) paper by Ned Block and Robert Stalnaker, "Conceptual Analysis and the Explanatory Gap, "http://www.nyu.edu/gsas/dept/philo/faculty /block/ExplanatoryGap.html--especially section 9. Nevertheless, in the present paper we suppose that Chalmers' approach is unproblematic and show that even so, his rejection of materialism does not follow.

[30] *The Conscious Mind*, p57.

[31] Putnam, "The Meaning of 'Meaning'," in K. Gunderson, ed., *Language, Mind & Knowledge* (Minneapolis: University of Minnesota Press, 1975), pp.131-193.

[32] *The Conscious Mind*, p.57.

[33] *Ibid*., p.59.

[34] *Ibid*., p.66.

is made on a priori grounds, he is most concerned with primary intensions. The first thing he must do is flesh out just how are we to make sense of "watery stuff". This is crucial. The primary intension of "water" is going to be something very close to a definite description.[35] The only codicil Chalmers makes is that we don't confuse the description which is fleshing out the intension with the function itself.

With the apparatus of primary and second intensions in place, it is a short step to two notions of possibility: 1-possibility and 2-possibility, one corresponding to each type of intension. Conceivability, then, is also divided into two classes: 1-conceivable situations and 2-conceivable situations. Since he is interested in our ability to reject materialism on a priori grounds, he needs to connect 1-conceivability to 1-possibility. A situation is 1-conceivable on Chalmers' view if it is conceivable according to the primary intensions of the terms involved. And, if something is 1-conceivable, it is 1-logically possible.[36] (From here on out, when we use the word "conceivable," we will mean 1-conceivable, since this is notion that Chalmers relies on for his argument.)

We now argue that given a Kripkean view of worlds and possibilia, zombie twins are impossible. What makes our objection compelling is that it falls out of the Chalmerian picture of primary intensions.

Remember that a possible world inhabited by your zombie twin is a counterfactual situation in which you lack phenomenal experiences. This seems readily conceivable to most people, especially philosophers. But this initial intuition is not strong enough for Chalmers. He needs it to be 1-conceivable, and thus 1-logically possible. So we must turn to the primary intensions of the concepts involved in a situation in which we are not conscious. In so doing, however, we find that our zombie twins are not 1-logically possible. Note that because Chalmers prizes the epistemic asymmetry of conscious phenomena, each of us will be describing a different (purportedly) 1-possible situation: Chalmers will be checking the 1-possibility of a situation in which Chalmers is not conscious, Dietrich will be checking one in which he is not conscious, Gillies will be doing the same for a situation in which he lacks phenomenal experiences, and so on. So in each of these situations there is an indexical concept: "Chalmers" in the situation he is imagining, "Dietrich" and

---

[35] *Ibid.*, pp.57-8.

[36] Chalmers' definition for a conceivable sentence is one that is true at all conceivable worlds (*Ibid.*, p.66). If he means this as it is written, giving sufficient conditions for a sentence's conceivability, then his examples in the text of what sentences and worlds are inconceivable amount to the fallacy of denying the consequent. If, on the other hand, he really means this quick definition as necessary and sufficient for a sentence being conceivable (so that his examples in the text are coherent), it still won't do for it has a consequence that all contingently true sentences are inconceivable! For the sake of charitability, we will regard '1-conceivability' as conceivability under the primary intensions of the terms involved, leaving 'conceivable' in its untutored sense.

"Gillies" in the appropriate cases, etc.

From what we know about primary intensions it is unproblematic to say that the primary intension of "Chalmers" is "chalmersy stuff", that of "Dietrich" is "dietrichy stuff" and similarly for "Gillies". But what is chalmersy stuff, anyway? Saying "Chalmers is chalmersy stuff" does not provide enough information about the primary intension of this concept to evaluate the possibility of the situations at hand. In order to get a finer-grained analysis we must turn to something like definite descriptions, taking care not to confuse the description with the function itself. Notice that such descriptions are only relevant to what the primary intension of "Chalmers" is *for Chalmers* (and likewise for Dietrich, Gillies, and so on), because what we want to know is whether the subject can imagine a counterfactual situation about herself in which she does not have conscious experience, while her psychological life remains unchanged.

Now, it seems reasonable that whatever the description is that Chalmers uses here to flesh out the primary intension of "Chalmers"—i.e. to give substance to "chalmersy stuff"—it must have consciousness as a constituent member. Thus he writes that "[o]ur core epistemic situation already includes conscious experience."[37] But then we have that conscious experience is part of the primary intension of "Chalmers", "Dietrich", "Gillies", etc., at least as evaluated by each of them. And if that is the case, then since all of this was to evaluate the logical possibility of counterfactual situations about ourselves, it follows that our zombie twins are not 1-possible. Hence, although zombie twins appear to be readily conceivable, given Chalmers' analysis of primary intensions, they are in fact not logically possible. So either the connection between conceivability and possibility that Chalmers crucially needs is not a tight one at all, or the connection is tight and (by modus tollens) zombie twins are not 1-conceivable (though we may think they are, in some rough sense of "conceivable").

(Note that even this much undermines Chalmers' reliance on conceivability. For the first horn of the dilemma, conceivability simply does not imply possibility. This means that, contra Chalmers, matters of what is or is not possible are *not* "accessible from the armchair."[38] For the second horn, conceivability does imply possibility, but we can't trust what we consider conceivable: though zombies seem 1-conceivable, they aren't, in fact. Matters of what is conceivable become very slippery. Again, armchair metaphysics is called into question.)

---

[37] *Ibid.*, p.193.

[38] *Ibid*, p. 68

Chalmers is likely to reply in one of several ways. First, he might insist that we can easily conceive (according to 1-conceivability) zombie twins of each other, and thus zombie twins are logically possible. A second approach might be to concede the impossibility of zombie twins, but maintain the 1-possibility of zombies in general. Finally, he might object that the situation calls not for the primary intension of "Chalmers" but for that of "physical Chalmers"—i.e., all the physical properties of Chalmers. We will address each of these in turn.

That we can conceive of Chalmers' zombie twin, and that he can conceive of ours, is of no consequence. The most salient way to make this point is to note that as far as each of us is concerned, we in fact *do* inhabit a world in which everyone is a zombie—everyone except ourselves, that is. Changing the problem of consciousness to the problem of other minds will not save Chalmers' argument.

Nor will it do to concede the impossibility of zombie twins, and try instead to make the case from the possibility of zombies simpliciter. As soon as we shift the discussion away from physically identical creatures to physically similar creatures, then the whole notion of logical supervenience would need overhauling. Moreover, at the very least such a move would leave materialism utterly unscathed for the materialist would simply retort: "Those zombies are just not physically enough like us to have conscious experience; they're not physically identical to us, and so of course it is possible that they are not conscious." Besides, such a reply by Chalmers to our argument has the costly price of departing from the Kripkean view of possible worlds, and thus would force Chalmers to deal with transworld identification.

Finally, objecting that what needs evaluating is not the primary intension of "Chalmers" but rather that of "the physical make-up of Chalmers" begs the question against the materialist. Since Chalmers is imagining a counterfactual situation in which he has no phenomenal experiences, saying that he needs only to imagine the physical respects of Chalmers already assumes that he can separate the physical from the phenomenal.

Now, we in fact concede that zombie twins are 1-conceivable. We just deny that this means they are possible. So the question emerges: "Given that zombie twins aren't 1-possible, why are they 1-conceivable?" Or to ask it another way: "Why do zombie twins *seem* logically possible, even though they aren't?"

We think something quite interesting is going on here. The argument of this section has established that zombies are impossible, given the Chalmerian semantic framework. Yet we agree with Chalmers that they are nevertheless conceivable. So, at least for the case of consciousness, what is conceivable is not a reliable guide to what is possible. It appears that when it comes to consciousness, our modal intuitions falter.

So we have some evidence that our modal intuitions should not be trusted where consciousness is concerned. And we also have some evidence that materialism is true -- our anti-zombie argument. But if our modal intuitions aren't to be trusted, it is not clear how much weight we should place on our anti-zombie argument.

In the next section, we argue that the nature of consciousness itself both explains why zombie twins seem possible and prevents us from ever seeing that it logically supervenes on the physical. In other words, we are going to use the nature of consciousness itself to argue that the conceivability of zombie twins does not entail their genuine possibility.

## V. THE LIMITS OF OUR IMAGINATIONS

Chalmers sometimes talks as if taking consciousness seriously entails being a dualist. But can't materialists take consciousness seriously, too? Our plaint is that they can. Pretheoretically, there is something question begging against the materialist when "Yes" is given to the question "Would conscious experience be any different if it did logically supervene?" (Of course, on Chalmers' theory of naturalistic dualism, we can't even ask this question. But since the jury is still out on his theory, the question is fair.)

Chalmers' anti-materialist arguments all attempt to bolster what we call the *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. Many people have this intuition, and Chalmers relies on it and exercises it. Chalmers uses this pumped-up Cartesian intuition to make plausible the logical possibility of zombie twins. In this section, we shall argue that we would still have our Cartesian intuitions even if consciousness logically supervened on the physical. We think that careful analysis shows that with respect to the Cartesian intuitions many of us have, the situation where consciousness logically supervenes would be exactly the same as the situation if Chalmers' arguments were correct. In short, we will show that consciousness itself leads to our Cartesian intuitions.

Doing this will undermine our confidence in 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 we can successfully doubt that our ability to conceive of zombie twins entails that zombie twins are logically possible. As a consequence, Chalmers' main anti-materialist argument loses its crucial premise that there is a logically possible zombie twin world. Note that here we are not trying to show that this crucial premise is false; we did that in section IV. And, we are not trying to show

that zombie twins do not *seem* logically possible; they clearly do (at least to many philosophers). Rather we are trying to show that being conscious makes it seem as if there is a logically possible zombie twin world; we are trying to explain why zombies seem possible in terms of consciousness itself, independently of any materialistic or dualistic claims. If we are right, then given our anti-zombie argument in section IV, materialism is true, but it will never seem true in an intuitive, reductionist way. We will be lead to the truth of materialism not because we see *how* it could be true, but merely because we can see *that* it must be true.

To be begin, suppose it really is the case that consciousness logically supervenes on the physical. So zombie twins 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.

To get the reader's intuitions going our way, note that how things seem to us when we conceive of them, 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, for example, 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. Geometry presents a good example. Many mathematicians attempted to prove Euclid's parallel postulate from the other postulates. They could and did conceive that the parallel postulate was a theorem. Yet, the parallel postulate turned out to be logically independent of the other postulates (giving rise to nonEuclidean geometry). So just because something is conceivable or seems logically possible doesn't mean it is.

Now for our core argument. 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 because you are already conscious of the quale on which your consciousness supervenes.

Here is a more detailed version of the argument. Call the physical, neurological 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, because experiencing CPP presupposes that you are antecedently conscious, i.e. that the CPP has already done its job.

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 our consciousness supervene on CPP; we'd have to see CPP causing our consciousness. But we can't do this because seeing CPP is having a conscious experience (of CPP -- a brain process). We are, as it were, forever locked inside our consciousness. We can't see anything causing it -- we can't see our consciousness supervening on CPP -- because we can't step outside our consciousness and watch it come into existence via some causal chain.

We will now use a variant of Searle's famous Chinese Room Argument to pump the intuition we want the reader to have. For our variant, we will assume, contra Searle, that the room is conscious and that the room's consciousness logically supervenes on the person-in-the-room (this assumption is similar to what is called the system's reply to Searle's argument).

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 we said, conscious experiences themselves. The room might be able to infer, somehow, that what the person-in-the-room is doing results in its being conscious, but we can do that too, as we will now argue. That is, we shall now argue that we can plausibly infer, independently of our anti-zombie argument, that consciousness does logically supervene on the physical.

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. But this seems oddly insubstantial, so maybe there is something all these extrinsic relations relate. Question: what could that be? Answer: *intrinsicness*. There must be some intrinsic properties all these relations are relating. But consciousness, as Chalmers notes, just is a case of an intrinsic, nonrelational property.[39] Now it is reasonable to suppose that intrinsicness and extrinsicness are very tightly related somehow, maybe even logically related. Therefore, it seems reasonable that consciousness (an intrinsic property) is tightly tied to the physical (extrinsicness) -- maybe even logically tied. But, if our arguments above are correct, we can never directly experience this supervening relationship. And in fact, we can't even get any decent direct empirical evidence for it. We are lead to hypothesizing its existence by logic and induction alone.

## VI. CONCLUSION

On our view, consciousness is a physical phenomenon (it logically supervenes on the physical), but because consciousness is the kind of phenomenon that it is (and why this should be is a big mystery), we are forever prevented from understanding how it could supervene on the physical. So, because consciousness is what it is, we are led erroneously to whatever Cartesian intuitions we may have. We can't see the supervenience base so we are lead to conclude there isn't one. It is quite natural to feel that our conscious experiences could be just what they are regardless of how the world is.

Our argument has an interesting mirror image quality when compared to Chalmers' theory of naturalistic dualism. Chalmers believes that the reason philosophers can conceive of zombie twins is because zombie twins are in fact logically possible[40] . He then uses the logical possibility of zombie twins and the fact that we are conscious to argue against materialism and for dualism. We believe, on the other hand, that zombie twins are logically impossible, but that we can conceive of them because we are conscious. The dissociation we feel between our consciousness and the world -- what we have called our Cartesian intuition -- is, on our view, a property of consciousness itself. Furthermore, we can partially explain this dissociation, whereas Chalmers cannot. Indeed, on his theory of naturalistic dualism, he is holding out for one day discovering the processes on which consciousness

---

[39] *The Conscious Mind,* p.153

[40] One could worry about the implicit circularity here in Chalmers' arguments: zombies are conceivable because they are possible, and possible because they are conceivable. Chalmers is aware of this pitfall and actually begins his argument with a weaker assumption: that we merely take consciousness seriously (xii). Unfortunately, it is not clear what this comes to.

naturally supervenes.

In sum, we agree with Chalmers that we can conceive of a zombie twin world. But even being very careful, we 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, Cartesian intuitions exist in all conscious, intelligent creatures. We can conceive of Chalmers being right about zombie twins, and we can conceive of Chalmers being wrong about zombie twins because we can conceive of mistakenly inferring the existence of zombies merely by the existence of our logically supervening consciousness. 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 we had to pick which conceivable world to go with, we'd pick the world where consciousness does logically supervene, but by its very intrinsic nature, prevents us from seeing this fact.