ON THE HARD PROBLEM OF CONSCIOUSNESS: AN EVALUATION AND DEFENSE OF

CHALMERS' TWO-DIMENSIONAL ARGUMENT AGAINST MATERIALISM

by

JAMES MARTIN GRINDELAND

(Under the Direction of René Jagnow)

ABSTRACT

I evaluate and defend David Chalmers' two-dimensional argument against materialism in light of criticisms. I motivate Chalmers' argument by showing how the traditional anti-materialist arguments of René Descartes, Saul Kripke, and Frank Jackson have serious conceivability problems. Specifically, they take certain statements to be conceivable, which might not really be conceivable; and they take conceivability as a reliable guide to possibility, despite the lesson of Kripkean a posteriori necessities that conceivability is not a perfectly reliable guide to possibility. Subsequently, I explicate Chalmers' argument and evaluate it in light of criticisms. I argue that there is a way of running Chalmers' argument that stands up well against the criticisms, and I conclude by summing up the most important ways in which my discussion advances the literature.

INDEX WORDS: conceivability, consciousness, materialism, necessity, physics, possibility,
                 Russellian monism, structure, two-dimensional semantics

ON THE HARD PROBLEM OF CONSCIOUSNESS: AN EVALUATION AND DEFENSE OF

CHALMERS' TWO-DIMENSIONAL ARGUMENT AGAINST MATERIALISM

by

JAMES MARTIN GRINDELAND

B.A., Minnesota State University Moorhead, 2004

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

ON THE HARD PROBLEM OF CONSCIOUSNESS: AN EVALUATION AND DEFENSE OF

CHALMERS' TWO-DIMENSIONAL ARGUMENT AGAINST MATERIALISM

by

JAMES MARTIN GRINDELAND

Major Professor: René Jagnow

Committee: Yuri Balashov
Charles Cross

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2013

DEDICATION

As in so many cases when a man is said to accomplish something, there is a woman behind the scenes helping to make it possible. In my case it was my wife, Thea, who took jobs outside of her field to accommodate our family needs, and did by far the greater amount of domestic work, including child rearing. We have raised our daughter, Estelle, with little money, in cramped housing, far away from relatives, all while accumulating massive student loan debt so that I can pursue a doctorate in philosophy—a field where the teaching job market is dire. Nevertheless my wife, Thea, always encouraged and supported the option of completing the degree despite the difficulties involved, and she pushed me hard to work towards completion. Thus I owe a great debt of gratitude to my wife, Thea, for her generous support and sacrifices that helped make this possible. We trust that it will pay off in ways that coincide with our values, so that, e.g., we can live a good life with our pride and joy, Estelle.

I would like to thank our daughter, Estelle, for putting so many smiles on our faces, and for helping me to put things in proper perspective. It has made me a better person.

Finally, I would like to thank my parents and grandparents for encouraging me to pursue higher education, and for supporting my decision to specialize in philosophy. My parents, who each earned graduate degrees while raising me, have been especially helpful. As a child, I absorbed academic ways of thinking from their example. While I was a graduate student, we had many phone conversations, and I benefitted greatly from their encouragement and advice. Moreover, my parents have always been very involved in my life and supportive of my interests, which has given me the confidence to take on difficult projects.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Science has an excellent track record of discovery. Because of painstaking scientific research and development, we have a great understanding of our environment. We have created telephones, cameras, video games, the internet, books, and hand-held wireless devices that incorporate all of this and more. We have measured the universe to be roughly 93 billion light years (10^{26} x 9.3 meters) wide, and we have measured its parts to be as small as the Planck length (10^{-35} meters). We understand ourselves well, too. Human birth and death is understood. Heredity is understood. Doctors routinely perform successful surgeries on us, including brain surgeries. Nowadays we are even learning how to grow body parts in the lab. There is little doubt that, provided we don't go extinct, scientists will discover pretty much everything.

One of the only remaining mysteries of the human being is why the correlations hold between our brain processes (e.g., C-fibers firing) and our conscious processes (e.g., pain). This "mind-body problem" hasn't been solved—yet.

Most problems, however, get solved in a scientific way, and so there is reason to expect the same in this case. Naturally, then, it may seem that the wisest attitude to have about this mind-body problem is to refrain from armchair ways of trying to solve it—for, say, a couple of hundred years. This would give scientists more time to figure out why the mind-body correlations hold. If they haven't figured it out by then, maybe then it will be worthwhile to try to figure it out from the armchair. Until then, our time is probably better spent in other ways—for instance, installing solar panels.

Yet, despite the plausibility of the thesis that we should temporarily refrain from armchair ways of trying to solve the mind-body problem, David Chalmers has given a promising two-dimensional argument against materialism that is *not* anti-scientific and that (I argue) constitutes *progress* on the mind-body problem. Indeed, his argument seems more plausible than the anti-materialist arguments of René Descartes, Saul Kripke, and Frank Jackson, and his conclusion suggests that solving the mind-body problem will require doing physics in a new way. Thus Chalmers' argument should be of interest for both philosophers and physicists or brain scientists.

The focus of this dissertation is to evaluate and defend Chalmers' two-dimensional argument against materialism in light of criticisms. I will start by motivating and outlining Chalmers' argument, then I will consider objections, and finally I will sum up the ways in which my contributions advance the literature.

In chapter 2, I set the stage for Chalmers' argument by arguing that the traditional anti-materialist arguments of René Descartes (1641a/2009, 1641b/2009), Saul Kripke (1971, 1972/2001), and Frank Jackson (1982, 1986) all have serious conceivability problems that render them inconclusive. Their conceivability problems come in the form of contentiously saying either that a given statement is conceivable, or that a given kind of conceivability entails metaphysical possibility. The latter kind of conceivability problem is especially serious, given the threat of naively overlooking Kripkean a posteriori necessities.

In chapter 3, I explain Chalmers' two-dimensional argument against materialism and its potential weaknesses. In this chapter, I also show that, although Chalmers' argument is constructed with the aim of minimizing conceivability problems, conceivability problems are still a serious threat for his argument. In chapters 4 and 5, I discuss these conceivability problems

in detail and show how Chalmers can deal with such problems. There are also problems related to the definition of materialism, which I discuss in chapter 6.

In chapter 4, I discuss criticisms of Chalmers' conceivability-to-possibility principles. These principles are supposed to warrant inferences from a certain kind of conceivability to a certain kind of possibility, and they are supposed to be more plausible than the conceivability-to-possibility principles employed by Descartes (1641a/2009, 1641b/2009), Kripke (1971, 1972/2001), and Jackson (1982, 1986). As we will see, counterexamples to Chalmers' conceivability-to-possibility principles imply what Chalmers calls "strong necessities," which are supposed to require that some logical possibilities are metaphysical impossibilities. I argue that Chalmers and others are wrong to think that the existence of strong necessities requires that some logical possibilities are metaphysical impossibilities. After this, I discuss a number of objections to Chalmers' conceivability-to-possibility principles. Namely, I discuss the objection that there may be mathematical strong necessities, the objection that there may be strong necessities clouded by epistemic vagueness, Stephen Yablo's (1999) necessary god argument, Anand Vaidya's (2008) "essential origins" argument, and Robert Howell's (2008) "reductio" argument. I argue that the objection from mathematical strong necessities is highly controversial, and that the other objections are unsound. Finally, I argue that Chalmers' argument should not employ the conceivability-to-possibility principle that is written in terms of *ideal negative conceivability*, and should instead employ the conceivability-to-possibility principle that is written in terms of *ideal positive conceivability*.

In chapter 5, I discuss the other sort of conceivability problem for Chalmers' argument. Namely, Chalmers contentiously asserts that such-and-such is conceivable. Roughly, he claims that it is conceivable (without cognitive defect) that a world be structurally indiscernible from

our world, yet *not* have all of our world's conscious properties. I discuss two criticisms: Stephan Leuenberger's (2008) "blocker" argument, and Keith Frankish's (2007) "anti-zombie" argument. I argue that Leuenberger's argument is unsound, and that Frankish's argument is invalid or unsound, depending on how it is formulated.

In chapter 6, I discuss objections related to the definition of materialism. To this end, I discuss different ways of defining materialism, and then some problems. The first problems are John Hawthorne's (2002) "blocker" argument and Jaegwon Kim's (1987/1993, 1989/1993, 1990/1993) "lone ammonium molecule" argument. According to these arguments, Chalmers' definition of materialism doesn't capture the right materialist commitments, in which case Chalmers' two-dimensional argument against materialism is misguided by virtue of discussing an uninteresting kind of materialism. I will also discuss Hempelian problems, which trace back to Carl Hempel (1969). According to these Hempelian arguments, a priori investigation of whether materialism is true is futile. Yet I will argue that Hawthorne and Kim's arguments are misguided, and that the Hempelian problems are not problems for Chalmers' two-dimensional argument against materialism.

Chapter 7 is the conclusion. Here I sum up the ways in which my discussion advances the literature, and I provisionally conclude that Chalmers' argument is sound and therefore that solving the mind-body problem will require a new kind of empirical discovery.

CHAPTER 2

TRADITIONAL MATERIALISM DEBATE

In this chapter, I set the stage for an evaluation and defense of David Chalmers' two-dimensional argument against materialism. This involves explaining the mind-body problem, giving a brief survey of the standard reasons for being a materialist, and explaining the traditional anti-materialist arguments. Particularly, I explain René Descartes' conceivability argument for mind-body dualism, Saul Kripke's argument against mind-brain type identity theory, and Frank Jackson's knowledge argument against materialism. I argue that these traditional anti-materialist arguments have serious conceivability problems that render them inconclusive. By contrast, in later chapters I will argue that these conceivability problems do not bedevil Chalmers' argument.

<u>Mind-body problem</u>

Let us adopt the conventional terminology that was independently coined by Timothy Sprigge in 1971 and later by Thomas Nagel in 1974 according to which for a creature to be *conscious* is for there to be "something that it is like" to be that creature.[1] Similarly, for a mental state to be a conscious state is for there to be something that it is like to be in that state. Examples of conscious states include the felt experiences of pains, colors, or an artist's visualization of a shape prior to drawing a representation of it. In the case of a conscious experiential feeling of pain, we are talking about the feeling itself, not mere dispositions to flinch, cry, or engage in

---

[1] Although Thomas Nagel (1974) is more famous for using the "something it is like to be" terminology, he points out in his 1986 book, *The View From Nowhere*, that, "two years earlier [i.e., two years before Nagel's (1974) paper was written] Timothy Sprigge (Sprigge 1971, p. 166-8) proposed as the essential condition of consciousness that there must be 'something it is like to be' the creature in question" (Nagel 1986, p. 15); yet Nagel clarifies that he did not plagiarize Sprigge; he says that he "hadn't read Sprigge" (Nagel 1986, p. 15).

escape behaviors. In the case of a conscious experience of color, we are not talking about certain wavelengths of the electromagnetic spectrum, such as a blind person may learn about, but rather about the color experiences that can be enjoyed by, say, those who appreciate fine art photography. And so it goes in general: to say that a state is a conscious state is to say that it has an experiential component. Moreover, it is clear that there *are* conscious experiences; for I have them.[2] I will also assume for the sake of argument that there is, roughly, a physical world, and that I have a brain. Given these assumptions, the question arises: How exactly are my conscious experiences related to the physical world, or to my brain? This is the "mind-body problem," "mind-brain problem," or "consciousness-physics problem;" although for convenience I will call it the *mind-body problem*.

Many philosophers have discussed the mind-body problem. For instance John Searle asks, "What is the relation of our minds to the rest of the universe? . . . How, for example, *could* this grey and white gook inside my skull be *conscious*? . . . this . . . oatmeal-textured substance of the brain."[3] Colin McGinn asks: "How can Technicolor phenomenology arise from soggy gray matter? . . . Somehow, we feel, the water of the physical brain is turned into the wine of consciousness, but we draw a total blank on the nature of this conversion. Neural transmissions just seem like the wrong kind of materials with which to bring consciousness into the world."[4] Douglas Hofstadter asks: "How can a self come out of stuff that is as selfless as a stone or a puddle? What is an "I," and why are such things found (at least so far) only in association with . . . certain kinds of gooey lumps . . . ?"[5] Stephen Kosslyn asks: "How can an idea arise from wet

---

[2] Dennett (1988; 1991) comes close to holding that there are no conscious experiences ("qualia"); yet he ultimately only denies their existence after packing controversial assumptions into the definition.
[3] Searle 1984, p. 14-23.
[4] McGinn 1989, p. 349; see also McGinn 1991, p. 1.
[5] Hofstadter 1999, p. P-2.

stuff?"[6] Thomas Huxley goes so far as to compare the proposal that "a state of consciousness comes about as a result of irritating nervous tissue" with rubbing a lamp to elicit a genie: "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the Djinn, when Aladdin rubbed his lamp."[7] So far, these descriptions are fairly crude, whereas the brain is a delicate, complicated system, with roughly 20-50 billion neurons and 100-200 trillion synapses. Philosophers understand this; sometimes they dumb down their characterizations for ease of reading, while other times they give more strict formulations. David Chalmers, for instance, calls the mind-body problem "the hard problem of consciousness,"[8] or, for short, "the hard problem."[9] "The hard problem of consciousness," he says, "is that of explaining how and why physical processes give rise to phenomenal consciousness."[10] He elaborates:

> It is undeniable that some organisms are subjects of experience, but the question of why it is that these systems are subjects of experience is perplexing. Why is it that when our cognitive systems engage in visual and auditory information processing, we have visual or auditory experience: the quality of deep blue, the sensation of middle C? How can we explain why there is something it is like to entertain a mental image or experience an emotion? It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner mental life at all? It seems objectively unreasonable that it should, and yet it does. . . . If any problem qualifies as *the* problem of consciousness, it is this one.[11]

This label, "the hard problem," has become standard in the literature. By calling it *the* hard problem, Chalmers doesn't mean to suggest that there are no other difficult problems about consciousness; it is just his rhetorical way of saying that, of all the problems having to do with consciousness, it is the most difficult that we're interested in. The hard problem results from

---

[6] Kosslyn 2003, p. 132.
[7] Huxley and Youmans 1868, p. 178.
[8] Chalmers 2002b/2010, p. 105.
[9] Chalmers 1995, p. 3.
[10] Chalmers 2002b/2010, p. 105.
[11] Chalmers 1995, p. 5.

looking for *explanations* of why we have the particular conscious experiences that we do, and

Chalmers says that it seems "objectively unreasonable" that "physical processing [should] give

rise to a rich inner mental life." In this way, Chalmers exhibits his anti-materialist tendencies.

Curiously, right after saying that it seems "objectively unreasonable" that physical

processing should give rise to conscious experiences, Chalmers follows this up by saying, "and

yet it does." The "and yet it does" clause is misleading, as it makes it seem as though *he* believes

that physical processing gives rise to consciousness. Yet the truth is that he is suspicious of this

view; at any rate, he is most sympathetic to the view that our world's fundamental reality is more

like conscious experience than is ordinarily supposed.[12] So, when Chalmers says, "and yet it

does," this appears to be a rhetorical way of saying that many philosophers have good reasons

for thinking that physical processing gives rise to consciousness. The mind-body problem is

therefore conducive to *cognitive dissonance*: for it seems both highly reasonable yet "objectively

unreasonable" to believe that physical processing gives rise to our conscious experiences.[13] Thus

the mind-body problem has many philosophers confused, puzzled, or baffled: Kripke says, "I

regard the mind-body problem as . . . extremely confusing;"[14] Sprague says, "The MIND-

BRAIN PROBLEM is a puzzler;"[15] and Chalmers says, "The mind-body problem is as baffling

as it ever was."[16]

To understand why many philosophers experience cognitive dissonance with respect to

the mind-body problem, it is important to understand the standard reasons for being pulled in

---

[12] More on this later, but for now note that Chalmers is very sympathetic to a view that he calls "panprotopsychism," which is an off-shoot of panpsychism, the view that consciousness is ubiquitous.

[13] This echoes Thomas Nagel's 1974 paper, "What is it like to be a bat?" and his 1986 book, *The View From Nowhere*. In those works he argues that, although we have excellent reasons to think materialism is true, we cannot understand how it could be so, since we lack a conceptual framework whereby something could be both publicly objective and privately subjective. In the absence of such a conceptual framework, the materialist thesis is bound to seem, as Chalmers puts it, "objectively unreasonable."

[14] Kripke 1972/2001, footnote 77.

[15] Sprague 1966, p. 71.

[16] Chalmers 1996, p. 25.

each direction. Thus in the next section I will outline the standard reasons for being a materialist;

and in the section after that I will discuss the traditional anti-materialist arguments.

Reasons for being a materialist

The standard reasons for being a materialist are as follows. First, brain scans (e.g., MRI, CT,

PET, EEG, MEG, NIRS) correlate brain activities with mental activities, including *conscious*

mental activities. Although correlations are consistent with many different explanations for why

the correlations hold, it is evidence that every aspect of our mentality is grounded in the brain.

Second, Occam's razor is a reason to think that our conscious experiences are identical to brain

processes, as postulating the existence of both brains and spiritual substances would arguably

multiply entities beyond necessity.[17] Third, Darwinian evolution is a reason to think that

conscious experiences are material phenomena; for it would seem strange if, after hundreds of

millions of years of material evolution, all of a sudden non-material phenomena would arise, not

from DNA, but coincidentally alongside complex brains that arose from DNA.[18] A fourth reason

is the success of science's methodological naturalism: many phenomena that have been assumed

to be non-material have turned out to have material explanations after all. Finally, a fifth type of

reason for being a materialist is on the basis of causal arguments. On the one hand, there are

"analytic functionalists" who identify mental phenomena with material causes as a definitional

maneuver, according to which it follows that mental phenomena are material. Yet there are also

identity theorists such as David Papineau who accept the existence of a material causal network,

and accept that mental phenomena have physical effects, but nevertheless think that it would be

absurd if every such effect was caused "twice over" by a material and a distinct mental cause; to

avoid this absurdity, they conclude that mental processes are *identical* to material processes and

---

[17] E.g., Smart 1959, p. 142; see also Place 1956.
[18] E.g., Lycan 2012, section 1.4, defends this view; although he doesn't specifically mention DNA.

therefore that they are not distinct causes.[19] Yet, despite the reasons for being a materialist, many philosophers are sympathetic to the traditional anti-materialist arguments.

Traditional anti-materialist arguments

The traditional anti-materialist arguments infer metaphysical possibility from conceivability. Soon we will see this as a problem, but for now let us simply recall such inferences as follows.

First, Descartes gives a conceivability argument for substance dualism. He starts by defining "thought" as "everything that is in us in such a way that we are immediately aware of it,"[20] and defining "mind" as "that substance in which thought immediately resides."[21] He defines "body" as "that substance which is the immediate subject of local extension and of the accidents that presuppose extension, such as shape, position, movement from place to place, and so on,"[22] and he defines "really distinct substances" as follows: "Two substances are said to be really distinct from one another when each can exist without the other."[23]

Given these definitions, Descartes investigates whether mind and body are *really distinct substances*. To be really distinct substances, each must be able to exist without the other. And so he tries to clearly conceive of each as existing without the other, and in this he takes himself to succeed. According to Descartes, he clearly conceives of his mind as existing with all of the thoughts that he in fact has, yet without having a body; likewise, he clearly conceives of his body as existing with all of the material properties that he in fact allegedly has, yet without having any thoughts; that is, that he is "merely an extended thing and not a thinking thing."[24] Finally, since

---

[19] See Papineau 2002, p. 17.
[20] Descartes 1641b, "Arguments Proving the Existence of God and the Distinction of the Soul from the Body, Arranged in Geometrical Fashion," Definition I, p. 72.
[21] Ibid, Definition VI, p. 72.
[22] Ibid, Definition VII, p. 72-3.
[23] Ibid, Definition X, p. 73.
[24] Nowadays philosophers would call such a being a "philosophical zombie," or, for short, a "zombie."

he clearly conceives of each existing without the other, he concludes that it is possible for each to exist without the other, and therefore that they are really distinct substances.

Descartes tries to make the argument even more intuitive with a religious articulation, suggesting that God can bring about any state of affairs that is clearly conceivable.

> Whatever we clearly perceive can be brought about by God in precisely the way we perceive it. . . . But we clearly perceive the mind . . . apart from any extended substance . . . and . . . we clearly perceive the body apart from the mind (as everyone readily admits). Therefore, at least by the divine power, the mind can exist without the body, and the body without the mind.
>   Now certainly, substances that can exist one without the other are really distinct . . . . But the mind and body are substances . . . that can exist one without the other (as has just been proved). Therefore, the mind and the body are really distinct.[25]

The idea here is that, since it is plausible that God can bring about any state of affairs that is clearly conceivable, Descartes concludes that clear conceivability entails possibility. Thus the clear conceivability of his mind without his body (and vice versa) entails the metaphysical possibility of his mind without his body (and vice versa). In this way, Descartes infers metaphysical possibility from clear conceivability.[26]

Second, Saul Kripke (1971; 1972/2001) argues against *type* identity theory—the view that every mental property, *m*, is identical to a physical property, *p*, in the sense that, necessarily, all instances of *m* are instances of *p*, and vice versa.[27] The most commonly cited example in the literature is that pain = C-fiber stimulation. Accordingly, Kripke challenges type identity theory

---

[25] Ibid, Axioms or Common Notions, Proposition IV, Demonstration, p. 75.

[26] God might not need to play any role in the argument. At any rate, Descartes cautions that his practice of citing God in the argument is "not because some extraordinary power is required to achieve [the] separation [of mind and body]" (Descartes 1641b, Proposition IV, p. 75), and that "it [is not] of any importance what power it is that separates two things for us to know that they are really distinct" (Ibid.). Thus he thinks that his argument is supposed to be just as good even if he doesn't cite God's existence. Nevertheless, he does believe that God makes the entailment from clear conceivability to possibility more plausible. For he says his mind and body can be separated "at least by the divine power" (Ibid.), and elsewhere he says that they can be separated "at least by God" (Descartes 1641b, Meditation VI, p. 64).

[27] As Kripke says: "Some philosophers have accepted the identity of particular sensations with particular brain states while denying the possibility of identities between mental and physical types. I will concern myself primarily with the type-type identities, and the philosophers in question will thus be immune to much of the discussion" (Kripke 1972/2001, p. 144).

by arguing that pain ≠ C-fiber stimulation. Kripke's argument goes as follows. For starters, "pain" and "C-fiber stimulation" are rigid designators.[28] A rigid designator, in Kripke's terminology, is anything that "designates the same object in all possible worlds."[29] But Kripke argues that identity statements flanked by rigid designators are necessary, so that true identity statements flanked by rigid designators are necessarily true, while false identity statements flanked by rigid designators are necessarily false. The proof of this last claim—*the necessity of identity statements flanked by rigid designators*—goes as follows. Everything has the property of being necessarily identical to itself; so, by the indiscernibility of identicals, it follows that, for any x and y, since x has the property of being necessarily identical to x, then x=y only if y also has the property of being necessarily identical to x.[30] So, since "pain" and "C-fiber stimulation" are rigid designators, pain = C-fiber stimulation only if pain has the property of being necessarily identical to C-fiber stimulation; that is, pain = C-fiber stimulation only if □(pain = C-fiber stimulation).

Yet "□(pain = C-fiber stimulation)" strikes Kripke as too strong a claim. For he *conceives* (at any rate, he holds that it "would seem"[31]) that C-fiber stimulation could have been felt as "tickle, or as warmth, or as nothing."[32] Moreover, from this conception, Kripke infers that these apparent counterexamples to the type identity thesis are *metaphysical possible.* That is, from the conceivability of these apparent counterexamples, Kripke infers ◊(pain ≠ C-fiber stimulation), hence ¬□(pain = C-fiber stimulation), which negates the type identity theorist's claim that

---

[28] Kripke 1972/2001, p. 148-9.
[29] Kripke 1971, p. 172. See also Kripke 1972/2001, p. 48: "Let's call something a *rigid designator* if in every possible world it designates the same object."
[30] See Kripke 1971, first paragraph, for his complete articulation.
[31] Kripke 1972/2001, p. 153.
[32] Kripke 1972/2001, p. 154.

□(pain = C-fiber stimulation). In this way, Kripke infers metaphysical possibility from seemingness (i.e., conceivability).

Third, Frank Jackson's (1982; 1986) knowledge argument against materialism is typically described as involving a hypothetical scientist, Mary, who "is confined to a black-and-white room" where she learns "everything there is to know about the physical nature of the world" through "black-and-white books and through lectures relayed on black-and-white television."[33] Yet, as Jackson asks in his 1982 paper, "What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she learn anything or not?"[34] He answers this question by appeal to *seemingness*; he writes: "It seems just obvious that she will learn something about the world and our visual experience of it."[35] Indeed, Jackson appeals to seemingness in his 1986 paper as well; he writes: "It seems . . . that [pre-release] Mary does not know all there is to know."[36] Rather, when Mary "is let out" of the room and "sees her first ripe tomato,"[37] it seems that she will learn something "*about the experiences of others*."[38]

> She will realize that there was, all the time she was carrying out her laborious investigations into the neurophysiologies of others and into the functional roles of their internal states, something about these people she was quite unaware of. All along their experiences (or many of them, those got from tomatoes, the sky, . . .) had a feature conspicuous to them but until now hidden from her (in fact, not in logic).[39]

Mary learns this about other people's experiences despite already knowing the true physics of our world. Indeed, she learns this despite being a strong enough reasoner that her failure to deduce it from the true physics was not due to a lack of "logical acumen and persistence,"[40] but

---

[33] Jackson 1986, p. 291.
[34] Jackson 1982, p. 130.
[35] Jackson 1982, p. 130.
[36] Jackson 1986, p. 291.
[37] Jackson 1986, p. 292.
[38] Jackson 1986, p. 292.
[39] Jackson 1986, p. 292-3.
[40] Jackson 1986, p. 292. Note: Chalmers prefers to stipulate that Mary is an *ideal* reasoner so as to rule out the option that Mary's failure to deduce the relevant conscious truths is a result of cognitive deficiency.

was because such facts are not implied by physical facts, in which case physicalism (i.e., materialism) is false.

Note that first Jackson says that it *seems*[41] that Mary will learn something that implies that materialism (i.e., physicalism) is false; later he says that she *will* learn something that implies that materialism is false. Thus it appears that Jackson's knowledge argument is a conceivability argument whereby there is an inference from the seemingness (i.e., conceivability) of a counterexample to materialism to the metaphysical possibility of a counterexample to materialism. In this way, Jackson infers metaphysical possibility from conceivability.[42]

Conceivability problems

As we have seen, the traditional anti-materialist arguments given by Descartes, Kripke, and Jackson all suppose that something is conceivable and infer that it is possible.[43] There is good reason to see this as a demerit—partly because what *seems* conceivable might not be conceivable, and partly because conceivability might not *entail* possibility. These are the two kinds of conceivability problem, as I will use the phrase.

Some things are conceivable but impossible, or inconceivable but not impossible; either way implies that conceivability is not a perfectly reliable guide to possibility. One example comes from Kripke: he cites the case of people who think that it is conceivable for water to be something other than H_2O. For instance, they might think it is conceivable for water to be

---

[41] Jackson doesn't only run a knowledge argument about Mary; he also runs one about Fred, who sees more colors than we do. Yet he appeals to seemingness here as well; he writes: "What kind of experience does Fred have when he sees red_1 and red_2? What is the new colour or colours like? We would dearly like to know but do not; and it seems that no amount of physical information about Fred's brain and optical system tells us" (Jackson 1982, p. 129).

[42] At any rate, the knowledge argument can be run this way. The alternative would be to run the argument by straightaway asserting that Mary will learn something that implies that materialism is false, which, in the absence of any further considerations, would beg the question against materialism.

[43] Thomas Nagel's (1974, 1986) *subjective/objective argument* is sometimes considered alongside the classic anti-materialist arguments, but strictly speaking Nagel does not defend an anti-materialist position; he merely argues that we do not understand the thesis of materialism, since we lack a conceptual framework for understanding how our conscious experiences (subjective) might be physical (objective).

Putnamian XYZ.[44] Yet Kripke argues that "water is H_2O" is a necessary truth. The argument goes as follows. First, water *is* identical to H_2O, which we discovered empirically. Second, "water" and "H_2O" rigidly designate the same *natural kind*. Third, recall that Kripke argues that identity statements flanked by rigid designators are necessary. Thus "water is H_2O," given that it is a true identity statement flanked by rigid designators, is necessarily true. Nevertheless, some people think that it is conceivable for water to be something besides H_2O, even though this is not really possible. Thus conceivability does not entail possibility.

A second example comes from Antoine Arnould: in his critique of Descartes' *Meditations*, Arnould asks: "How does it follow, from the fact that he [Descartes] is aware of nothing else belonging to his essence [as a thinking thing], that nothing else [e.g., spatial extension] does in fact belong to it?"[45] Arnould then imagines a case where conceivability fails to coincide with possibility: namely, he says that we can imagine a person who thinks s/he can clearly conceive of a right-angled triangle that does not obey the Pythagorean theorem. Although such a triangle would be impossible, someone might be this naïve about triangles; at any rate, it wouldn't be the first time someone got a math problem wrong. So, this is another case proving that conceivability does not entail possibility.

A third example is that many philosophers throughout history have thought it inconceivable for there to be multiple physical entities occupying the same place at the same time.[46] Yet they are likely wrong, as physicists have held for nearly one hundred years that multiple bosons *can* occupy the same place at the same time, even while having all of the same

---

[44] See Putnam 1975/1985.
[45] Descartes 1641b, p. 143.
[46] To give just three examples: Locke (1690, p. 367) says it is "impossible for two things of the same kind to be or exist in the same instant, in the very same place;" Gale (1973, 413) says: "Two or more material objects of the same kind cannot wholly occupy the same place at the same time;" finally, Descartes ("Meditation 2," p. 44) packs it into his definition of "body," saying: "By 'body', I understand all that is capable of being bounded by some shape, or being enclosed in a place, and of filling up a space in such a way as to exclude any other body from it."

qualitative properties. This is a well-established, empirically informed belief; so, in so far as the science is right, it proves that inconceivability does not entail impossibility.

This brings us to a possible fix: we might say that only non-naive conceptions entail possibility, whereas confused or naïve conceptions do not entail possibility. We will see later that philosophers nowadays commonly upgrade to *ideal* conceivability to increase the plausibility of an entailment of possibility. Yet there's a trade-off lurking in the dialectic: namely, the less naive a conception must be in order to entail possibility, the more difficult it will be to achieve such a conception in the first place. Thus the more "ideal" one's conception is supposed to be, the more likely it will be that one's argument is unsound for the reason that one's conceivability premise is false. Likewise, relatively lax standards regarding how "ideal" one's conception must be can be met more easily, but without as much assurance that the corresponding possibility will be entailed by the conceivability. Confused or naïve conceptions don't carry as much modal weight.

Altogether, then, when we are trying to figure out what is possible, the use of *conceivability* as a philosophical tool should be seen as problematic. For this reason, many philosophers have given up hope of figuring out whether materialism is true. They suspect we are in no epistemic position to take a justified stand one way or another. The hard problem, they say, is probably too hard for humans to figure out;[47] maybe even too hard for any material being, or any possible philosopher.

Given the seriousness of the conceivability problems for the traditional anti-materialist arguments, these arguments are inconclusive. One might therefore be tempted to embrace the argument in favor of materialism explicated above. But, as we will see, a close consideration of Chalmers' argument will show that this conclusion would be premature.

---

[47] McGinn (1989) gives the classic defense of this position.

CHAPTER 3

CHALMERS' TWO-DIMENSIONAL ARGUMENT

In chapter 2, I argued that the traditional anti-materialist arguments of René Descartes, Saul Kripke, and Frank Jackson are bedeviled by conceivability problems that render them inconclusive. By contrast, in this chapter I explain and give a reconstruction of Chalmers' two-dimensional argument against materialism, which aims to avoid these conceivability problems. I also indicate potential weaknesses with Chalmers' argument. Highlighting these potential weaknesses will allow me to structure the rest of my dissertation, as the primary purpose of later chapters will be to evaluate Chalmers' argument in light of its potential weaknesses.

Chalmers' (1993) dissertation defends property dualism and rejects materialism. His (1996) book, *The Conscious Mind*, which is a re-working of his dissertation, also defends property dualism and rejects materialism. Yet in Chalmers' (2002a) paper, "Does Conceivability Entail Possibility?" he does neither. Instead of defending property dualism or rejecting materialism, he upgrades to the disjunctive conclusion that materialism is false or panprotopsychism is true. And, in more recent papers, he defends the disjunctive conclusion that materialism is false or Russellian monism is true. For instance, in his (2005) paper, "The Two-Dimensional Argument Against Materialism," he gives a "full" version of the argument. His (2009) paper is an abridged version of his (2005) paper. And his (2005/2010) paper is an updated version of the (2005) paper. It is this most recent formulation that I will tend to focus on.

Understanding Chalmers' arguments requires understanding Russellian monism. Russellian monism is the view that phenomenal properties are either fundamental features of the

world, or else they are constituted by protophenomenal properties that are not themselves phenomenal but merely constitute phenomenal properties when arranged in the right ways. Since 2005, Chalmers has concluded that materialism is false or Russellian monism is true. Prior to 2005, Chalmers stated his conclusion differently, and never used the phrase "Russellian monism." Thus, on the face of it, it might seem that Chalmers' argument has changed drastically over the years: namely, that he used to argue that materialism is false, but that now he argues that materialism is false or Russellian monism is true.

Upon close inspection, however, we can see that the philosophical changes are minimal and that the differences are largely terminological. For instance, in Chalmers' (1996) book, he has a different way of talking about Russellian monism. He defends property dualism,[48] which he says might "turn out to be a kind of monism,"[49] whereby "the physical and the phenomenal will turn out to be two different aspects of a single encompassing kind."[50] He says, "There are two ways this might go."[51] The first way is to "take experience itself as a fundamental feature of the world, alongside space-time, spin, charge, and the like."[52] The alternative way is to hold that "there is some *other* class of novel fundamental properties from which phenomenal properties are derived. . . . We could call these properties *protophenomenal* properties, as they are not themselves phenomenal but together they can yield the phenomenal."[53] Clearly, these two ways that property dualism might go are precisely the two ways that Russellian monism might go. Thus the Russellian monism that Chalmers discusses in recent years implies the property dualism

---

[48] Chalmers 1996, p. 124-9.
[49] Chalmers 1996, p. 129.
[50] Chalmers 1996, p. 129.
[51] Chalmers 1996, p. 126.
[52] Chalmers 1996, p. 126.
[53] Chalmers 1996, p. 126-7.

that he defends in his (1996) book. Thus one of the biggest differences between Chalmers' more recent arguments and his (1996) argument is terminological.

Russellian monism, to be clear, is named after Bertrand Russell's view that at least some events have phenomenal intrinsic natures, even though physics only describes them in the way that physics describes anything: mathematically. To be clear, Russell held that physics only specifies positions and movements without specifying any non-relational properties of the entities that have these positions or that move in these ways. Russell makes this point on many different occasions.[54] One of the most memorable formulations is in Russell's (1959/1997) book, *My Philosophical Development*, when he says,

> It is not always realised how exceedingly abstract is the information that theoretical physics has to give. It lays down certain fundamental equations which enable it to deal with the logical structure of events, while leaving it completely unknown what is the intrinsic character of the events that have the structure. . . . All that physics gives us is certain equations giving abstract properties of their changes. But as to what it is that changes, and what it changes from and to – as to this, physics is silent.[55]

Thus Russell says that physics describes our world's entities in an "exceedingly abstract" way, and that it merely describes "abstract properties of their changes": namely, their changing positions. As for non-relational features of these entities (i.e., their "intrinsic character"), "physics is silent." Elsewhere Russell says that physics merely describes "mathematical" properties, and he takes this as a sign that physics is *epistemically impoverished*: "Physics is mathematical, not because we know so much about the physical world, but because we know so little: it is only its mathematical properties that we can discover."[56]

Chalmers explains Russell's view in the following way, speaking as one who accepts the view:

---

[54] E.g., Russell 1921, 1924, 1927a, 1927b/2001, 1948, and 1959/1997.
[55] Russell 1959/1997, p. 13.
[56] Russell 1927b/2001, p. 123.

> Physical theory only characterizes its basic entities relationally, in terms of their causal and other relations to other entities. Basic particles, for instance, are largely characterized in terms of their propensity to interact with other particles. Their mass and charge is specified, to be sure, but all that a specification of mass ultimately comes to is a propensity to be accelerated in certain ways by forces, and so on. Each entity is characterized by its relations to other entities, and so on forever. . . . The picture of the physical world that this yields is that of a giant causal flux, but the picture tells us nothing about what all this causation relates. Reference to the proton is fixed as the thing that causes interactions of a certain kind, that combines in certain ways with other entities, and so on; but what is the thing that is doing the causing and combining? As Russell (1927[a]) notes, this is a matter about which physical theory is silent.[57]

Once again, Russell's idea is that physics only describes how our world's entities are arranged at a time, or over a span of time. At a time, entities have positions *relative* to one another. Over a span of time, entities have changing positions *relative* to each other. Yet these are the only kinds of descriptions that physics ever gives. Chalmers calls arrangements at a time "structure," and he calls changing structure "dynamics."[58] Thus it's part of the Russellian philosophy of physics that physics only describes structure and dynamics.

Russell is not alone in thinking that physics only describes structure and dynamics. For instance, Bas van Fraassen echoes the point, saying: "Science abstracts, it presents us with the structural skeleton of nature only."[59] The physicist Arthur Eddington writes in his (1920) book, *Space, Time, and Gravitation*: "Physics is the knowledge of structural form, and not knowledge of content.[60, 61] So, Russell, Chalmers, van Fraassen, and Eddington hold that physics describes nothing but structure and dynamics, which are mere relations.

By only describing relations, physics doesn't describe any non-relational properties. Yet Russell proposes that non-relational properties exist, and that some of the intrinsic natures are

---

[57] Chalmers 1996, p. 153.
[58] E.g., Chalmers 2002b/2010, p. 120.
[59] van Fraassen 2006, p. 6.
[60] Eddington 1920, p. 200.
[61] Russell might have even gotten the idea from Eddington's (1920) book. For the earliest I know of Russell making the point is in his (1921) book, *The Analysis of Mind*, whereas I know that Russell had read Eddington's (1920) book at least by 1926, as he discusses it in his (Russell 1926) article, "Philosophical Consequences of Relativity."

conscious experiences. Although (as far as I can tell) Russell never worked with a concept of a protophenomenal being or property, Chalmers stretches the phrase "Russellian monism" to cover that option as well. Thus Russellian monism—on Chalmers' terminology—holds that there are phenomenal or protophenomenal occurrences, but that physics only describes their structural and dynamical relations.

The last prerequisite before I sketch Chalmers' two-dimensional argument against materialism is to sketch a definition of materialism. Materialism is roughly the view that any physically indiscernible world has at least the positive facts of our world. Being *physically indiscernible* means having the *same physics*. Having the same physics means something different, depending on whether Russell, Chalmers, et al. are right that physics doesn't specify intrinsic natures. If physics doesn't specify intrinsic natures, two worlds can be physically indiscernible even if their physical events have different intrinsic natures. Yet, if physics does specify intrinsic natures, then being physically indiscernible does imply that their physical events have the same intrinsic natures.

As we will see, Chalmers' argument doesn't require taking a stand on whether physics specifies intrinsic natures. Rather, it argues that, if physics doesn't specify intrinsic natures, materialism is false; whereas, if physics does specify intrinsic natures, Russellian monism is true; either way, materialism is false or Russellian monism is true.

Preliminary sketch of Chalmers' argument

Chalmers' two-dimensional argument against materialism begins with a *Conceivability Premise*, which says, roughly, that it is ideally conceivable that there be a world that has our world's structure and dynamics but does not have all of our world's positive conscious truths. From here, he infers the corresponding *Possibility Premise*, which says, roughly, that it is possible for there

to be a world that has our world's structure and dynamics but that does not have all of our world's positive conscious truths. Chalmers takes this inference from conceivability to possibility to be warranted on the basis of a conceivability-to-possibility (CP) principle that I will soon discuss. Finally, Chalmers derives his conclusion from the *Possibility Premise* combined with the following *Tautologous Disjunction*: either

 (a) Our world's structure, dynamics, and intrinsic natures together *do not* entail our world's positive conscious truths; or else

 (b) Our world's structure, dynamics, and intrinsic natures together *do* entail our world's positive conscious truths.

Either option implies Chalmers' conclusion that materialism is false or Russellian monism is true. On the one hand, (a) straightforwardly implies that materialism is false; hence it implies that materialism is false or Russellian monism is true. Meanwhile (b) and the *Possibility Premise* together imply Russellian monism. For, if the *Possibility Premise* is right that our world's structure and dynamics do not entail all of our world's positive conscious truths, but (b) is right that our world's structure and dynamics combined with our world's intrinsic natures do entail all of our world's positive conscious truths, then there's a key role being played by our world's intrinsic natures in the entailment of our world's positive conscious truths, which requires that Russellian monism be true. Thus, regardless of whether (a) or (b) is true, Chalmers' conclusion follows: materialism is false or Russellian monism is true.

Rigorous formulation of Chalmers' argument

I will now give a rigorous formulation of Chalmers' two-dimensional argument against materialism. To begin with some terminology, he lets "P" stand for a conjunction of all of our world's microphysical truths, he lets "T" stand for a "that's all" clause which says that our world

doesn't have any fundamental truths besides those expressed by "P," he lets "Q" stand for an

arbitrary positive conscious truth in our world, and he puts atomic statements side-by-side to

represent their conjunction. For instance, XY&Z would be Chalmers' shorthand for (X&Y)&Z.

With this terminology out of the way, Chalmers outlines his two-dimensional argument against

materialism as follows:

(1) $PT\&\neg Q$ is conceivable.
(2) If $PT\&\neg Q$ is conceivable, $PT\&\neg Q$ is 1-possible.
(3) If $PT\&\neg Q$ is 1-possible, $PT\&\neg Q$ is 2-possible or Russellian monism is true.
(4) If $PT\&\neg Q$ is 2-possible, materialism is false.
(5) Materialism is false or Russellian monism is true.[62]

Line 1 is Chalmers' *Conceivability Premise*. In line 1, when Chalmers says that PT&¬Q is

conceivable, there are three implicit qualifications. First, he intends to be speaking of *ideal*

conceivability as opposed to mere *prima facie* conceivability; second, he intends to be speaking

of *positive* or *negative* conceivability; and third, he intends for the formula "PT&¬Q," to be

evaluated by *1-intensions*, which I will explain below.

First, in line 1 Chalmers implicitly means to be saying that PT&¬Q is *ideally*

conceivable as opposed to merely *prima facie* conceivable. An *ideal* conception would involve

no cognitive defect; or, as Chalmers puts it, "contingent cognitive limitation."[63] It cannot be

sloppy, confused, or epistemically improvable. By contrast, *prima facie* considerations are

merely initial considerations, which often involve oversights.

Since it is more difficult to have an ideal conception, Chalmers' *Conceivability Premise*

is that much more likely to be false. Yet if it is true when qualified to be about ideal

conceivability, this makes it that much more likely that what is thereby conceivable is also

possible. For an ideal conception has a better chance of entailing possibility than does a non-

---

[62] This is very close to Chalmers' outline in Chalmers 2010, p. 161.
[63] Chalmers 2010, p. 143.

ideal conception. For ideal conceptions, by definition, are not naïve about the natures of what they conceive. Thus there is no chance that an ideal conception will fail to entail possibility by virtue of naïvely conceiving of the nature of what is thereby conceived.

Of course, given the qualification of ideal conceivability, there is the following problem: since we are not ideal reasoners, how we can we know whether Chalmers' *Conceivability Premise* is true? After all, Chalmers' *Conceivability Premise* might seem true to us, but merely given our naivety. Chalmers argues, however, that, although we are not ideal reasoners, it is still possible for us to reason in an ideal way about certain things:

> Although we are nonideal, we can know that it is not ideally conceivable that 0=1 and that it is ideally conceivable that someone exists. We know that certain things about the world (say, that all philosophers are philosophers) are knowable a priori and that certain things about the world (say, that there is a table in this room) are not so knowable even by an ideal reasoner.[64]

Thus Chalmers gives several plausible examples of statements that we can know to be either ideally conceivable, or not ideally conceivable. Moreover, he holds that the same kind of reasoning makes his *Conceivability Premise* plausibly ideally conceivable, as he writes: "Likewise, reasoning of this sort gives us very good reason to think that there is no a priori entailment from physical to phenomenal truths and that zombie hypotheses are conceivable even for an ideal reasoner."[65]

Second, in line 1 Chalmers implicitly means to qualify PT&¬Q as being *positively* or *negatively* conceivable. (The person giving the argument is supposed to pick one or the other.) Chalmers describes positive conceivability as follows: "S is positively conceivable when one can coherently imagine a situation in which S is the case."[66] As defined, a positive conception of some hypothesis S involves coherently imagining a case in which S is the case. Here we can

---

[64] Chalmers 2010, p. 155.
[65] Chalmers 2010, p. 155.
[66] Chalmers 2010, p. 144.

quibble about what it would take to coherently imagine a situation in which S is the case, as different things might come to mind for different kinds of hypotheses. For physical hypotheses, it might seem natural for a relevant "situation" to be *a particle-by-particle description of a world*. There is a problem with this in general, however, which is that, for any necessary truth, such as a mathematical truth, this method won't work. For necessary truths are not contingent upon a particular distribution of physical particles. After all, any particle-by-particle description of a world would be a case where every mathematical truth is true, whereas it seems that no such conception can rightly be called a positive conception of every mathematical truth. Similarly, it would seem wrong to require that a positive conception of S must involve *a conscious mental image*. For a conscious mental image would involve a way of sensing certain contingent features that a world might have, none of which count as forming a positive conception of a necessary truth. So, Chalmers should take some heat for this. His notion of what it means to be a positive conception is not fleshed out. No algorithm is given to determine whether a given conception is positive.

Yet perhaps we could allow a disjunction of options, so that, e.g., a particle-by-particle description of a world might be sufficient to positively conceive of a world physically; a conscious mental image might be sufficient to positively conceive of sensory properties; and there might be other methods that would be sufficient for positively conceiving of mathematical truths, and so on.

Meanwhile Chalmers describes negative conceivability differently: "S is [ideally] negatively conceivable when S cannot be ruled out through a priori reasoning."[67] Thus, whereas the ideal positive conceivability of S is the possibility of ideally coherently imagining S, the ideal

---

[67] Chalmers 2010, p. 144.

negative conceivability of S is the necessity of failing to notice incoherence in S while trying (via an ideal reasoning process) to notice incoherence in S.

Chalmers doesn't take a stand on which qualification is better for the purpose of his *Conceivability Premise*. Ideal negative conceivability is arguably easier to achieve, but less likely to entail possibility. Ideal positive conceivability is arguably more difficult to achieve, but more likely to entail possibility. So, in so far as we want to make sure that what we claim to be conceivable really is conceivable, we should run the conceivability premise with negative conceivability; whereas, in so far as we want to make sure that the conceivability entails a corresponding possibility, we should run the conceivability premise with positive conceivability.

Third and finally, in line 1 Chalmers implicitly means to evaluate PT&¬Q by 1-(primary) intensions. 1-intensions are elements of Chalmers' two-dimensional semantics, which I will explain below. As we will see, 1-intensions allow us to specify our world's structure and dynamics, while 2-intensions allow us to specify our world's intrinsic natures. Chalmers appeals to both intensions, which is what renders his argument "two-dimensional."

Intensions are familiar in possible world semantics. An intension tells us what a given reference is in each possible world. An intension associated with the word "canine" tells us which things are canines in each possible world. Likewise, an intension associated with a statement tells us which possible worlds are the worlds where that statement is true, and which possible worlds are the worlds where that statement is false.

The relevant intensions for Chalmers' two-dimensional argument against materialism are 1- and 2-intensions. 1-intensions, which he calls "primary intensions," are defined in terms of concepts (which tend to be approximations of reality). 2-intensions, which he calls "secondary intensions," are defined in terms of the way reality is in its maximum degree of precision. Thus,

while 1-intensions are more closely associated with a person's concepts, 2-intensions tend to incorporate more detail about the true nature of reality than is captured by our fallible concepts.

For instance, if suddenly there is a masked intruder, my concept of this masked intruder is apt to be pretty general: roughly, *the masked intruder,* perhaps along with some general impressions about the height and build of the person. Meanwhile there is a lot of detail about the masked intruder that my concept leaves out. Maybe the masked intruder is a banker; maybe the masked intruder is a woman; maybe the masked intruder has such-and-such number of electrons; etc. Anyway, there is this aptness for difference in detail between 1- and 2-intensions, since 1-intensions are grounded in fallible concepts, which are often feeble attempts at comprehension, whereas 2-intensions are grounded in reality in its maximum degree of precision.

To give another example, consider "water." A 1-intension of *my* word "water" will map each possible world to a set that is only as fine-grained as *my* concept. Thus, if I am not well versed in chemistry, my concept might be pretty general: for instance, it might be roughly the set of all stuff that is more or less watery as far as I can tell by looking at it with a naked eye, feeling it with my hands, or tasting it, etc. For short, my concept of water might merely be that it is *watery stuff*. If so, the 1-intension of my word "water" will map each possible world to the set of all watery stuff; i.e., the set, $\{x \mid \text{watery}\}$. (This formula can be read, "the set of all x such that x is watery.") Meanwhile, the 2-intension of my word "water" will map each possible world to a more specific set: namely, the set of whatever our world's watery stuff *is* in its maximum degree of detail. Thus, if our world's watery stuff is H_2O (if that is its essence in its maximum degree of detail), then a 2-intension will map each possible world to the set, $\{x \mid H_2O\}$. As is typical,[68] the

---

[68] I say that this is *typically* how 2-intensions behave with respect to corresponding 1-intensions; although it is possible for them to behave differently. For instance, "9+4=13" is necessary regardless of whether it is evaluated by a 1- or a 2-intension. Thus here we can say that, with respect to the statement, "9+4=13," the 1- and 2-intensions

2-intension picks out a *proper subset* of the set that the 1-intension picks out. For {x | watery}

includes more than just H$_2$O, as it also includes Putnamian XYZ.

To give yet another example, consider "Hesperus" and "Phosphorus." Hesperus is "The

Evening Star," while Phosphorus is "The Morning Star." Yet 2-intensions of "Hesperus" and

"Phosphorus" map each possible world to the same planet: Venus. For Venus appears in a certain

part of the sky in the morning and in a certain part of the sky in the evening, and it was an

empirical discovery that each appearance was of the same astronomical body. For all we knew a

priori, there might have been distinct astronomical bodies, one appearing thusly in the morning,

the other appearing thusly in the evening. Thus as far as 1-intensions of "Hesperus" and

"Phosphorus" go, it is plausible that they pair up some possible worlds with distinct referents,

since it is plausible that some possible worlds are very nearly like ours but include the exception

that the astronomical bodies in the specified parts of the sky in the mornings and evenings are

distinct astronomical bodies.

To give a still further example, consider "lightning." A 2-intension of "lightning" maps

every possible world to a certain kind of electrical discharge, yet a 1-intension might map some

possible worlds to something other than electrical discharge. For instance, a 1-intension might

map worlds where Berkeleyan idealism is true to certain ideas, since in Berkeleyan worlds there

is no electricity *per se*, but only minds and their ideas.

We have now seen several examples where the mappings between 1- and 2-intensions

differ with respect to the specificity of their algorithms. They were all cases where the 2-

intension used a more specific "description" with which to pick out an extension. A 1-intension

picked out, roughly, *the masked intruder*; whereas the 2-intension more specifically picked out

coincide in the sense that, necessarily, their referents/extensions are the same. (Both map every possible world to the truth value, TRUE.)

every last detail of the intruder. A 1-intension picked out, roughly, *watery stuff*; whereas the 2-intension more specifically picked out H_2O. 1-intensions picked out, roughly, *the astronomical body that appears thusly in the morning, and the astronomical body that appears thusly in the evening*; whereas the corresponding 2-intensions more specifically picked out Venus. A 1-intension picked out, roughly, a *lightning-ish event* consistent with being mere Berkeleyan ideas, whereas the 2-intension more specifically picked out a kind of electrical discharge.

The same difference might hold with respect to physics statements. For we are arguably naïve about the intrinsic natures of our world's physics. All we understand about our world's physics, conceptually, is the structure and dynamics involved. We assign mass, charge, quantum wave function, etc.; but none of this describes our world's intrinsic natures in a non-structural/dynamical way. Thus it is plausible that physics statements, when evaluated by 1-intensions, only specify structure and dynamics, since that is the only sort of thing that we understand about the true physics.[69] Meanwhile, if the events described by physics have intrinsic natures, then our physics statements, when evaluated by 2-intensions, will specify those intrinsic natures as well. Thus, when we want to speak only about our world's structure and dynamics, we can do so by giving physics statements and specifying that they are to be evaluated by 1-intensions, whereas when we want to also speak of our world's intrinsic natures, we can do so by giving physics statements and specifying that they are to be evaluated by 2-intensions.

---

[69] Here there is an implicit assumption that there is a true physics. To deny this would be to go beyond the usual anti-realist philosophies of science. For instance, Quine (1953/1980) argues that any statement, including any scientific theory, should be considered tentative and revisable in principle. However, this is a claim about our social practice of epistemology, not a metaphysical thesis about the nature of our world. Similarly, van Fraassen (1980) defends a "constructive empiricist" view about what science aims to do: he claims that science aims for "empirical adequacy" but that its claims about unobservables need not be literally true. However, his claim is about what science aims to do, not about the way that the world is. Finally, those who accept the verificationist theory of meaning (e.g., Ayer 1952) are committed to holding that statements about unobservables are meaningless since their truth cannot be verified empirically. Strictly speaking, this does not contradict the assumption that there is a true physics; it merely puts a constraint on what a true physics might be like: it implies that a true physics cannot incorporate unobservables, since, if it did, the "claims" about unobservables would be meaningless and thus not true.

In sum, Chalmers' *Conceivability Premise* has three implicit qualifications. The conceivability is supposed to be ideal conceivability, positive[70] conceivability, and "PT&¬Q" is supposed to be evaluated via 1-intensions. As such, we can re-write Chalmers' *Conceivability Premise* in such a way as to make these qualifications explicit. The only problem is that we don't yet know how to specify that PT&¬Q is to be evaluated by 1-intensions.

If Chalmers were to say that PT&¬Q, when evaluated by 1-intensions, is possible, he would notate this by saying that PT&¬Q is 1-possible. Thus one would expect that, if Chalmers were to say that PT&¬Q, when evaluated by 1-intensions, is conceivable, he would notate this by saying that PT&¬Q is 1-conceivable. But he didn't say this in line 1; he merely said that PT&¬Q is conceivable. It is not clear why he didn't say that it was 1-conceivable, as there appears to be no advantage in keeping the qualification implicit. However, I argue that either way of notating the relevant intensions—i.e., "1-possibility" or "1-conceivability"—is problematic. For, either way, a given intension is applied to the entire statement in question. Yet there are plenty of cases where we will want to evaluate some parts of a given statement by a 1-intension, while evaluating other parts of the same statement by a 2-intension. For instance, we might want to say that it is possible for water to not be water, but to evaluate the first instance of "water" by a 1-intension while evaluating the second instance of water by a 2-intension. Chalmers should want this statement to come out true, since he believes that not all possible water in the first sense (watery stuff) is water in the second sense (H_2O). Unfortunately, Chalmers notation isn't capable of formulating this statement.

---

[70] For convenience, I will say "positive" and keep the "negative" option as an implicit alternative. For remember that the argument can be run either way. As Chalmers puts it: "The argument forms that follow can be understood as generating two different arguments, depending on whether one understands conceivability as ideal primary positive conceivability or as ideal primary negative conceivability" (Chalmers 2010, p. 148).

Fortunately, there is an easy fix. Namely, for each subformula that is supposed to be evaluated it by a particular intension, simply put a numerical subscript at the end of it denoting the intension by which to evaluate it. Put numeral 1 subscripts at the end of the subformulas that are to be evaluated by 1-intensions, and put numeral 2 subscripts at the end of the subformulas that are to be evaluated by 2-intensions by the rest. For instance, then, the claim about water that Chalmers' notation isn't capable of making can be stated as follows: "It is possible for $water_1$ to not be $water_2$." This is a straightforward fix, and is easily understandable.

Moreover, then, when we incorporate Chalmers' implicit qualifications of ideal conceivability, positive conceivability, and this new way of indicating which intensions with which to evaluate various subformulas, his two-dimensional argument against materialism can be re-stated as follows:

1′. $P_1T_1 \& \neg Q_1$ is ideally positively conceivable.

2′.   If $P_1T_1 \& \neg Q_1$ is ideally positively conceivable, $P_1T_1 \& \neg Q_1$ is possible.

3′.   If $P_1T_1 \& \neg Q_1$ is possible, $P_{1,2}T_{1,2} \& \neg Q_{1,2}$ is possible or Russellian monism is true.

4′.   If $P_{1,2}T_{1,2} \& \neg Q_{1,2}$ is possible, materialism is false.

5.   Materialism is false or Russellian monism is true.

In line 1′ (the reformulation of Chalmers' *Conceivability Premise*), P_1T_1 says that the conceived scenario has the same structural/dynamical profile that obtains in our world, $w_@$. For P_1 guarantees that the conceived scenario has every positive structural/dynamical truth as obtains in $w_@$, while the "that's all" clause, T_1, guarantees that it doesn't have any positive structural/dynamical truths that do not obtain in $w_@$. Finally, $\neg Q_1$ says that the positive conscious truth, Q_1, does not obtain in the conceived scenario. Altogether, line 1′ says that it is ideally positively conceivable for there to be the same structure and dynamics as in our world,

$w_@$, yet where Q_1 does not obtain. Instead of instantiating Q_1's truthmaker, such a scenario would either instantiate a different conscious property in its place (thereby involving a *phenomenal inversion*), or no conscious property at all in its place (thereby involving a *philosophical zombie*).

Line 2′ says that $P_1T_1\&\neg Q_1$ is ideally positively conceivable only if it is also possible. Chalmers justifies this with the following conceivability-to-possibility principle:[71]

CP+   Ideal primary positive conceivability entails primary possibility.[72]

The idea here is that, necessarily, for any statement S, S_1 is ideally positively conceivable only if $\Diamond S_1$. So, since it is true of any statement, it is true of $P_1T_1\&\neg Q_1$.

By modus ponens, lines 1′ and 2′ imply the following *Possibility Premise*: $\Diamond(P_1T_1\&\neg Q_1)$. Thus line 2′ can be read as claiming that the *Conceivability Premise* implies the *Possibility Premise*. The *Possibility Premise* is not an explicit premise in Chalmers' argument, but it is what you get if you execute a modus ponens operation on lines 1′ and 2′.

At this stage in the argument, Chalmers believes that his most implausible premises have been stated. As he puts it: "The real work in the argument is done by the first and second premises."[73] This is right, although let us see how to complete the argument.

Line 3′ says that $P_1T_1\&\neg Q_1$ is possible only if either $P_{1,2}T_{1,2}\&\neg Q_{1,2}$ is possible or Russellian monism is true. As we will see, this is stronger than necessary. For one thing, as Chalmers points out, his argument doesn't require Q to be evaluated by a 2-intension; it merely requires Q to be evaluated by a 1-intension. Moreover—this is something Chalmers does not say—evaluating T by a 2-intension is redundant. For, as long as T_1 says that there are no other

---

[71] If Chalmers were to run the argument with *negative* conceivability, he would cite CP- instead, which says, "Ideal primary negative conceivability entails primary possibility" (Chalmers 2010, p. 147).
[72] Chalmers 2010, p. 147.
[73] Chalmers 2010, p. 143.

basic positive structural-dynamical truths, this will ensure that there aren't any other intrinsic natures either, since, if there were other intrinsic natures, they would add additional structure. Thus we can re-write line 3′ as follows:

3″.   If $P_1T_1\&\neg Q_1$ is possible, $P_{1,2}T_1\&\neg Q_1$ is possible or Russellian monism is true.

Chalmers never formulates his third premise in this way, and perhaps this is partly because his way of notating intensions isn't flexible enough to specify that different subformulas are to be evaluated by different intensions. Fortunately, however, he does explain the justification for line 3″. He explains it by pointing out what it would take for the possibility of $P_1T_1\&\neg Q_1$ to *not* imply the possibility of $P_{1,2}T_1\&\neg Q_1$. Namely, if the former is true but the latter is false, this is to say that our world's structure and dynamics (and *that's all* the structure and dynamics) does not entail Q_1, but that our world's structure, dynamics, and intrinsic natures (and *that's all* the structure, dynamics, and intrinsic natures) does entail Q_1. This would imply Russellian monism. For if our world's structure and dynamics do not entail Q_1, but our world's structure, dynamics, and intrinsic natures do entail Q_1, then our world's intrinsic natures play a key role in the entailment of Q_1. As Chalmers puts it: "This idea—that the structural properties of physics in the actual world do not necessitate the existence and/or nature of consciousness but that the intrinsic properties of physics combined with the structural properties do—corresponds to a familiar view in the metaphysics of consciousness. This view is the type-F monism of the last chapter, or Russellian monism."[74] Thus the only alternative to the claim that $\lozenge(P_1T_1\&\neg Q_1)$ implies $\lozenge(P_{1,2}T_1\&\neg Q_1)$ is the Russellian monism option according to which our world's intrinsic natures play a key role in the entailment of the positive conscious truth, Q_1. Thus line 3″ is true.[75]

---

[74] Chalmers 2010, p. 151.
[75] For Chalmers' explanation of the justification, see Chalmers 2010, p. 149-151.

Line 4′ says that $\Diamond(P_{1,2}T_{1,2}\&\neg Q_{1,2})$ implies that materialism is false. As before, T and $\neg Q$ do not need to be evaluated by 2-intensions; so we can re-write line 4′ as follows:

4″.   If $P_{1,2}T_1\&\neg Q_1$ is possible, materialism is false.

Deriving line 4″ requires deriving a sufficient condition for materialism's falsity. For line 4″ says that the possibility of $P_{1,2}T_1\&\neg Q_1$ is sufficient for materialism's falsity.

We can see that the possibility of $P_{1,2}T_1\&\neg Q_1$ is sufficient for materialism's falsity. For materialism is roughly the view that our world's microphysics entails all of our world's positive truths, including all of our world's positive conscious truths. Consequently, if materialism is true, then any world that has the same structure, dynamics, and intrinsic natures as our world will have every positive truth that our world has. To put this into symbols, let "A" stand for a conjunction of all of our world's positive truths. Then the prose formula above can be re-written as follows:

M_x Materialism is true of $w_@$ only if $\Box(P_{1,2}T_1 \supset A_{1,2})$.

The fact that "P" and "A" are to be evaluated by a 2-intension will ensure that their intrinsic natures (if there are any) are taken into account. This is good because we are trying to say that, if materialism is true, then any world that has the same structure, dynamics, *and intrinsic natures* will have every positive truth that our world has. Having every positive truth that our world has will include having Q_1, as A_1 includes Q_1 as one of its conjuncts. Thus M_x implies the following:

$M_@$ Materialism is true of $w_@$ only if $\Box(P_{1,2}T_1 \supset Q_1)$.[76]

By contraposition on $M_@$, $\neg\Box(P_{1,2}T_1 \supset Q_1)$ implies that materialism is false of $w_@$. Equivalently, $\Diamond(P_{1,2}T_1\&\neg Q_1)$ implies that materialism is false of $w_@$. That is, line 4″ is true.

---

[76] Chalmers does not use a symbol to stand for a conjunction of *all* of our world's positive truths, whereas I use "A" for this purpose. However, it is important to have a symbol for this purpose; otherwise we won't be able to derive that materialism is true of our world only if $\Box(P_{1,2}T_1 \supset Q_1)$.

Finally, we can derive Chalmers' conclusion. For by modus ponens on the *Possibility Premise* and line 3″, we can derive the following:

3.1. $P_{1,2}T_1 \& \neg Q_1$ is possible or Russellian monism is true.

According to line 4″, line 3.1's left disjunct implies that materialism is false. So, line 3.1's left disjunct implies that materialism is false or Russellian monism is true. Likewise, line 3.1's right disjunct also implies that materialism is false or Russellian monism is true. Either way, line 5 is true. So, line 5 is true: *materialism is false or Russellian monism is true*.

<u>Strengths and weaknesses</u>

Whereas the traditional anti-materialist arguments of Descartes, Kripke, and Frank Jackson have serious conceivability problems that I outlined in chapter 2, Chalmers incorporates three modifications to help minimize such problems. As we have seen, he qualifies the conceivability as ideal, positive (or negative), and primary.

First, qualifying the conception as *ideal* makes the inference from the *Conceivability Premise* to the *Possibility Premise* more plausible. For, by qualifying the conceivability as ideal, this implies that it could be conceived without cognitive defect or relevant naivety. This is important: for, if it could only be conceived with cognitive defect or relevant naivety, such conceivability would be a poor reason to think it's possible. For, if it could only be conceived with cognitive defect or relevant naivety, the most natural explanation of this is that the thing that is said to be conceivable doesn't quite make sense, which would be evidence that it isn't possible. Thus incorporating the qualification of *ideal* in the *Conceivability Premise* comes with the following trade-off: it decreases the plausibility that the *Conceivability Premise* will be true, but it increases the plausibility that the *Conceivability Premise* implies the *Possibility Premise*. Is the trade-off worth it? It definitely is. For, although it somewhat decreases the plausibility that

the *Conceivability Premise* will be true, it is important to include the qualification of ideal, since the inference to the *Possibility Premise* would be more or less hopeless otherwise. Meanwhile, by including the qualification of ideal, it renders the inference from the *Conceivability Premise* to the *Possibility Premise* far from hopeless, and indeed quite plausible.

Second, Chalmers' argument can be run with either positive or negative conceivability. Each way has advantages and disadvantages. Since it is more difficult to achieve a positive conception than to achieve a negative conception, qualifying the conceivability as *positive* renders the *Conceivability Premise* less plausible. For it is less plausible that someone can positively conceive of $P_1 T_1 \& \neg Q_1$. Yet achieving a positive conception is arguably a better indicator that what is conceived is also possible; for its coherence is in some sense "pictured" or "imaged" or "grasped" in the conception itself. Thus, when the argument is run with positive conceivability, the inference to the *Possibility Premise* should be that much more plausible. Again, though, the argument can be run either way, with positive or negative conceivability. In this way there is flexibility, so that one can run the argument either way depending on whether one wants to strengthen the plausibility of the *Conceivability Premise* or the plausibility that the *Conceivability Premise* implies the *Possibility Premise*.

Third, qualifying the conception as *primary* makes the *Possibility Premise* more plausible than if it had been written in terms of both 1- and 2-intensions. That is, it is more plausible that $P_1 T_1 \& \neg Q_1$ is possible than that $P_{1,2} T_1 \& \neg Q_1$ is possible. This is because of the theoretical possibility that P_1 is consistent with multiple different intrinsic natures as the realizer of the structure and dynamics that P_1 asserts. The idea is that the more different possible arrangements of intrinsic natures that all have the same structure and dynamics, P_1, the better the chance that at least one of those arrangements is such that *it* (and that's all) does not entail Q_1. So, since "P_1" is

consistent with the option that alternative arrangements of intrinsic natures could be at play, it is more plausible that P_1T_1 might hold in a world where Q_1 is false. Suppose I say that Willie Nelson's *guitar* can be used to play Stella by Starlight without sounding like a guitar. That is not very plausible. But if I say that Willie Nelson has an *instrument* that can be used to play Stella by Starlight without sounding like a guitar, that is more plausible, since he might have, say, a piano.

The fundamental point is that, as a rule of thumb, we should accept the following *odds of failure* principle:

> OF   The more options there are for what might fail to φ, the more likely it is that at
>
> least one of them fails to φ.

We can apply this "odds of failure" principle, OF, to Chalmers' argument as follows. $P_{1,2}T_1$ is consistent only with the option of P_2 as the intrinsic natures, whereas P_1T_1 is consistent with multiple different options of intrinsic natures: P_2, $P_{2'}$, $P_{2''}$, etc. Thus, for an arbitrary φ, we should think that, given $P_1T_1 \& \varphi$ and OF, $\lozenge(P_1T_1 \& \neg\varphi)$ is more likely than $\lozenge(P_{1,2}T_1 \& \neg\varphi)$. Thus, substituting Q_1 for φ, we get the result that, given $P_1T_1 \& \neg Q_1$ and OF, $\lozenge(P_1T_1 \& \neg Q_1)$ is more likely than $\lozenge(P_{1,2}T_1 \& \neg Q_1)$. This makes the *Possibility Premise*, $\lozenge(P_1T_1 \& \neg Q_1)$, more plausible than if it had been written as $\lozenge(P_{1,2}T_1 \& \neg Q_1)$. It makes the *Conceivability Premise* more plausible for the same reason. For an ideal reasoner would consider OF, and apply it to the case.

Another strength of Chalmers' two-dimensional argument is that it does not require assuming that Q_1 and Q_2 are equivalent. Many philosophers think Q_1 and Q_2 are equivalent on the grounds that there is no appearance/reality distinction when it comes to conscious experiences. Kripke, for instance, argues that there is no difference between the way pain feels and the way that pain is in itself, given that "pain" is *defined* as the way it seems or feels; as he

puts it, "pain" rigidly designates the feeling of pain.[77] Similarly, he thinks that *every* conscious experience is defined as the way it seems or feels. John Searle agrees, saying: "*Where appearance is concerned we cannot make the appearance-reality distinction because the appearance is the reality.*"[78] Chalmers, too, is sympathetic with this line of thought, saying:

> It is plausibly the case that anything that feels like consciousness is consciousness. So it is not clear that the notion of "pseudoconsciousness," something that satisfies the primary intension of 'consciousness' without being consciousness, is coherent. The same holds for other, more specific phenomenal properties. So there is a strong case that the primary and secondary intensions of phenomenal terms coincide.[79, 80]

Thus several philosophers, including Chalmers, believe it is plausible that Q_1 and Q_2 are equivalent. Nevertheless it is possible to argue that they are not equivalent. For instance, one might hold that, whereas Q_1 denotes the presence of phenomenal consciousness, Q_2 denotes a token intrinsic nature, which might not be the only possible token intrinsic nature that can be correlated with Q_1. However, those who hold that Chalmers' argument requires that Q_1 and Q_2 be equivalent,[81] despite Chalmers' insistence to the contrary,[82] are mistaken. For we can easily see in the formulation above that Chalmers' argument doesn't require this, as it never even uses the term "Q_2." My notation makes clear that Chalmers' argument doesn't need to assume that Q_1 and Q_2 are equivalent. For my notation is able to say "$P_{1,2}T_1 \supset Q_1$;" his notation wasn't able to say this, as his notation can only say that an entire statement is to be evaluated via a 1-intension, or that an entire statement is to be evaluated via a 2-intension.

---

[77] Kripke 1972/2001.
[78] Searle 1992, p. 122.
[79] Chalmers 2010, p. 150.
[80] To ward off a potential misunderstanding, note that, strictly speaking, Chalmers doesn't hold that we are *incorrigible* with respect to our beliefs about which phenomenal properties we have; his point is just that there is no difference between how they are presented to us in experience and how they are *really* presented to us in experience.
[81] E.g., Kallestrup 2006, p. 282-3; Brüntrup 2011, p. 19.
[82] Chalmers writes: "it is worth noting (contrary to a common supposition) the assumption that Q has the same primary and secondary intensions is not necessary for the argument to go through" (Chalmers 2010, p. 153).

Yet Chalmers' two-dimensional argument against materialism has a number of potential weaknesses. We can safely ignore the inferences that are uncontroversially truth-preserving. Yet, aside from these, one potential weakness is Chalmers' *Conceivability Premise*:

1′. $P_1T_1\&\neg Q_1$ is ideally positively conceivable.

Some philosophers reject this premise. In chapter 5, I discuss two such arguments. The first is from Stephan Leuenberger (2008): he argues that line 1′ is false, as he thinks that we can never form a positive conception of something that incorporates a "that's all" clause. He argues that this would require positively conceiving the absences of sufficiently unfamiliar properties. I argue that Leuenberger's argument is unsound. The second objection is Keith Frankish's (2007) "anti-zombie" argument. It is a parity of reasons argument to the conclusion that Chalmers' *Conceivability Premise* is false. I argue that it is invalid, and that, if modified so as to be valid, the modified formulation will be unsound; so that, either way, it is not a good parity of reasons argument against Chalmers' two-dimensional argument against materialism.

Another potential weakness is line 2′:

2′. If $P_1T_1\&\neg Q_1$ is ideally positively conceivable, then $P_1T_1\&\neg Q_1$ is possible.

This says that the *Conceivability Premise* implies the *Possibility Premise*. Chalmers cites the conceivability-to-possibility principle, CP+, as its justification. (Again, if we were to run the argument in terms of negative conceivability, we are supposed to cite CP- as the justification.) In chapter 4, I explain Chalmers' defense of his CP principles. I also clarify what it would take for there to be a counterexample to his CP principles, and then I discuss objections to his CP principles. First, I discuss an objection to CP- according to which there may be mathematical necessities whose negations are ideally negatively conceivable. I argue that this objection is highly contentious. Second, I discuss an objection to CP- according to which there are necessities

that are unknowable due to the epistemic vagueness of the corresponding language. I argue that

such a necessity would absurdly require a logical necessity's necessity to be logically arbitrary,

hence logically contingent, hence not logically necessary, and that the epistemic theory of

vagueness is furthermore implausible by virtue of requiring a problematic philosophy of

reference. Third, I discuss an objection according to which both the existence and non-existence

of a necessarily existing God is ideally conceivable, and thus (by CP) such a God is absurdly

both possible and impossible. I argue that such a God is not ideally conceivable after all. Fourth,

Anand Vaidya (2008) argues that CP+ is false. He starts by rejecting the view that necessity

implies necessary necessity, and he uses this to help argue that some logical possibilities are

metaphysical impossibilities, and he uses this to help argue that CP+ is false. I argue that

Vaidya's argument is unsound. Fifth, Robert Howell (2008) argues roughly that, given our

ability to conceive of there being coherent conceptions that are not logically possible, CP, if true,

is false. Moreover, he says that a "two-dimensionalist" can resist this argument only by

undermining two-dimensionalism in other ways. In reply, I argue that Howell mischaracterizes

Chalmers' argument, and that his critique is implausible.

A third class of potential weaknesses for Chalmers' two-dimensional argument against

materialism has to do with the definition of materialism. Chalmers assumes that materialism is

true of $w_@$ only if $\Box(P_{1,2}T_1 \supset A_{1,2})$. This assumption is employed in the derivation of line 4′. Yet it

is not obvious that it can be derived from a *tenable* definition of materialism. In chapter 6, I

discuss two challenges. First, Hawthorne (2002) argues that the sorts of definition of materialism

that Chalmers favors are problematic in that they implicitly make controversial assumptions

about what is impossible. I argue that Hawthorne's argument is unsound. Second, Kim

(1987/1993, 1989/1993, 1990/1993) argues that the sorts of definitions of materialism that

Chalmers favors are written in terms of global supervenience, whereas intuitively they should be written in terms of local supervenience. I argue that Kim's objection is both weak and irrelevant to the plausibility of Chalmers' two-dimensional argument. Furthermore, in addition to these objections, there is what is supposed to be a more insidious problem related to the definition of materialism: a Hempelian critique whose basic lesson is that more scientific discovery is needed before we can have a more productive discussion about whether materialism is true. I argue that this is not a problem for Chalmers' two-dimensional argument.

CHAPTER 4

ARE CHALMERS' CP PRINCIPLES TRUE?

Depending on whether Chalmers runs his argument with positive or negative conceivability, he infers the *Possibility Premise* from the *Conceivability Premise* by appeal to one of the following conceivability-to-possibility principles:

CP+   Ideal primary positive conceivability entails primary possibility.

CP-   Ideal primary negative conceivability entails primary possibility.

Unlike the conceivability-to-possibility principles that are employed (explicitly or implicitly) in the traditional anti-materialist arguments of René Descartes, Saul Kripke, or Frank Jackson, CP+ and CP- are more sophisticated. In this chapter, I will first outline Chalmers' reasons for believing his CP principles. Then I will argue that there is a mistake in Chalmers' defense of his CP principles. Finally, I will discuss a number of objections to his CP principles: first, an objection to CP- according to which there may be mathematical necessities whose negations are ideally negatively conceivable; second, an objection to CP- according to which there are necessities that are unknowable (and thus whose negations might be ideally negatively conceivable) due to the epistemic vagueness of the corresponding language; third, an objection according to which both the existence and non-existence of a necessarily existing God is ideally conceivable; fourth, Anand Vaidya's (2008) "essential origins" argument against CP+; and fifth, Robert Howell's (2008) reductio argument against Chalmers' "metaphysical two-dimensionalism."

<u>Chalmers' defense of CP+ and CP-</u>

Chalmers defends CP+ and CP- on the grounds that there can be no counterexamples to them. A

counterexample would be a statement, S, such that S_1 is ideally conceivable but impossible

(equivalently, $\Box\neg S_1$). The necessity, $\Box\neg S_1$, would not be knowable a priori (since S_1 is ideally

conceivable), and thus it would be knowable (if at all) only a posteriori. Thus any

counterexample to CP+ or CP- would imply the existence of a statement that, when evaluated by

1-intensions, is necessarily true, yet only knowable (if at all) a posteriori. Chalmers calls these

"strong a posteriori necessities;" or, for short, "strong necessities."[83] As he puts it,

> A strong necessity . . . is an *a posteriori* necessity with a necessary primary intension.
> (This should be taken as the defining property of a strong necessity.)[84]

Chalmers argues that strong necessities are impossible. For he thinks that their possibility would

(absurdly) require *modal dualism*:

> MD   Every metaphysical possibility is logically possible, but not every logical
>
> possibility is metaphysically possible.[85]

According to MD, the space of metaphysical possibility is in some sense "smaller" than the

space of logical possibility. For every metaphysically possible world is a logically possible

world, but not every logically possible world is a metaphysically possible world. Yet it's still not

clear why the possibility of strong necessities should require MD. Chalmers aims to explain the

connection as follows:

> Advocates of strong necessities . . . use such modal notions as consistency, rational
> entailment, and conceivability themselves (their position is partly defined with respect to
> these notions). So they must accept something akin to the space of logically possible
> worlds, although they might use another name. But they think there is a further

---

[83] See Chalmers 2010, p. 167.
[84] Chalmers 1999, § 3.1.
[85] It is a virtue of this characterization of modal dualism that it doesn't presuppose that the collection of all logical or metaphysical possibilities forms a set. This makes the formulation less contentious, as some philosophers (e.g., Pruss 2001) argue that such collections do not form sets.

metaphysical modality, and that not every logically possible world is a metaphysically possible world. This modality is not to be defined in terms of the rational modality, or even in terms of the rational modality plus nonmodal facts; it is a further primitive, part of a different circle entirely.

This picture is *modal dualism*. It requires two modal primitives: there is a space of logically possible worlds, and then (as a further primitive fact) a smaller space of these picked out as "metaphysically possible". The primitives need not be cast in these terms, but we know that two primitives will be needed, as we need one to account for the rational modalities, and we need another in turn to account for the "metaphysical".[86]

Chalmers' idea here is that strong necessities are defined partly in terms of ideal conceivability, since they are defined as having ideally conceivable negations. Thus advocates of strong necessities are thereby committed to "such modal notions as consistency, rational entailment, and conceivability." These notions imply a space of rational possibility, "something akin to the space of logically possible worlds." Since this space of rational possibility is entailed by the defining notions of ideal conceivability, ideal conceivability entails rational possibility. But, by the definition of strong necessities, being ideally conceivable (thus rationally possible) doesn't entail being metaphysically possible, since the negations of strong necessities are ideally conceivable (hence rationally possible) but not metaphysically possible. Thus strong necessities require a kind of modal dualism. Assuming the space of rational possibility is *identical* to (not merely *akin* to) the space of logical possibility, we can we can say that strong necessities require MD.

Yet Chalmers argues that MD is absurd and strong necessities are impossible. In *The Conscious Mind* he gives six reasons for this,[87] and in his (2010) paper he says that he still accepts most of these reasons, although he doesn't say which ones.[88]

Chalmers' most powerful reason to reject MD is that it seems "brute and inexplicable" why a given logically possible world should not also be metaphysically possible. As he puts it, "It may be reasonable to countenance brute, inexplicable facts about *our* world, but the existence

---

[86] Chalmers 1999, § 3.5.
[87] Chalmers 1996, p. 137-8.
[88] Chalmers 2010, p. 170.

of such facts about the space of possible worlds would be quite bizarre. The realm of the possible (as opposed to the realm of the natural) has no room for this sort of arbitrary constraint."[89] Thus Chalmers acknowledges that *our world* may contain "brute, inexplicable facts." And this much is plausible.[90] But then he continues, saying, "the realm of the possible (as opposed to the realm of the natural) has no room for this sort of arbitrary constraint."

The only hint about why Chalmers thinks the realm of the possible has no room for this sort of arbitrary constraint comes in the form of an example about philosophical zombies—for short, zombies. Zombies are defined as being physically identical to conscious people except that they are not conscious. Many philosophers hold that zombies are logically but not metaphysically possible; yet Chalmers uses the metaphor of "God" to convey how unjustified this latter proposal should seem:

> Someone who holds that a zombie world is logically possible but metaphysically impossible has to answer the key question: *Why couldn't God have created a zombie world?* Presumably it is in God's powers, when creating the world, to do anything that is logically possible. Yet the advocate of metaphysical necessity must say either the possibility is coherent, but God could not have created it, or God could have created it, but it is nevertheless metaphysically impossible. The first is quite unjustified, and the second is entirely arbitrary.[91]

There are two options: (a) "the possibility is coherent, but God could not have created it," or (b) "God could have created it, but it is nevertheless metaphysically impossible." Chalmers says that option (a) seems "quite unjustified." This seems right. For option (a) says that "the possibility is

---

[89] Chalmers 1996, p. 137.

[90] If our world had no brute, inexplicable facts, every fact would have a Leibnizean sufficient reason, which appears to have the absurd result that every fact is necessary. At any rate, Peter van Inwagen (van Inwagen 1983, p. 202-4) argues as follows. Suppose there are some contingencies. Let C stand for a conjunction of all the contingent truths, and assume the principle of sufficient reason. Since C is a conjunction of contingent truths, C is contingent. But then there's a dilemma: namely, the sufficient reason for C is either necessary or contingent. If the sufficient reason for C is necessary, C is necessary, contrary to the fact that it is contingent. Meanwhile, if the sufficient reason for C is contingent, it is contained in C, in which case it cannot be a sufficient reason for C, since nothing can be a sufficient reason for itself. So, on this latter horn, there is and isn't a sufficient reason for C. So, both horns are contradictory. So, at least one of the two starting assumptions—that there are some contingencies, or that the principle of sufficient reason is true—is false. So, if the principle of sufficient reason is true, there are no contingencies, i.e., all truths are necessary, which is an apparently absurd result.

[91] Chalmers 1996, p. 138.

coherent, but God could not have created it." There doesn't seem to be any good reason to believe this. Why think that some coherent possibilities are metaphysically impossible? For, besides incoherence, there doesn't seem to be anything that could render something metaphysically impossible. Likewise, Chalmers seems right that option (b) is "entirely arbitrary." For it seems that the only way to consistently hold that zombies are metaphysically impossible, despite the fact that God could have created them, is to call them "metaphysically impossible" as a mere act of linguistic fiat. Yet, in so far as the space of metaphysically possible worlds is constrained by mere linguistic fiat, it is just an irrelevant distraction from substantive philosophical issues. So, option (a) is "quite unjustified," while option (b) is "entirely arbitrary." Either way is absurd, and since any statement that is logically but not metaphysically possible would involve analogous absurdity, MD is absurd as well.

Moreover, even if we set option (b)'s arbitrariness aside, Chalmers points out that materialism is false anyway if God can create zombie worlds that are metaphysically impossible. As he puts it, "after fixing the physical facts about the world, God still had more work to do."[92] For, if God were to try to create a zombie world, he could start by creating the physics. At this point, it would be a zombie world, but it wouldn't yet be metaphysically possible. The physics alone would not entail all of the positive facts that would be required in order to transform it into a metaphysically possible world. Thus, in addition to the fact that options (a) and (b) are absurd, option (b) is also a case according to which materialism is false, so that it is misguided (in an ironic way) to appeal to option (b) in an objection to an anti-materialist argument such as Chalmers' two-dimensional argument against materialism.

Besides his argument that MD is absurd, Chalmers gives some other reasons to reject MD as well. For instance, he holds that MD is fairly unmotivated. As he puts it: "The only real

---

[92] Chalmers 1996, p. 138.

46

motivation for this view would seem to be to save materialism at all costs, perhaps because of perceived problems with dualism."[93] Indeed, he points out that there is an irony here as well, which is that, while such philosophers are trying to be good naturalists, "the invocation of brute 'metaphysically necessary' principles constraining the space of possible worlds introduces an element much more problematic, and indeed far less naturalistic, then the mere invocation of further natural laws postulated by property dualism."[94] Thus the irony is that they are trying to defend a naturalistic position with an argument that postulates an extremely non-natural fact. Admittedly, this is not a knock-down objection to MD. For being both weakly and hypocritically defended is consistent with being true.

Additionally, Chalmers finds it troubling that, if MD were true, his 1- and 2-intensions would have to be evaluated over both spaces of possible worlds. Thus every statement would have a 1-intension that's evaluated over the logically possible worlds, an intension similar to the 1-intension that's evaluated over the metaphysically possible worlds, a 2-intension that's evaluated over the logically possible worlds, and an intension similar to the 2-intension that's evaluated over the metaphysically possible worlds. Moreover, the collection of logically possible worlds and the collection of metaphysically possible worlds would be distinct, not only from each other, but also from the collection of "naturally possible worlds" (i.e., the nomically possible worlds; i.e., the collection of metaphysically possible worlds that are consistent with our world's laws of nature), so that there would ultimately be *three* spaces of worlds, and two intensions to be evaluated over each. The realm of possibility would be *complicated*. Yet this isn't a knock-down objection to MD either, as the problem of "being complicated" isn't very serious, given that complexity is consistent with truth. It might be *inconvenient* to keep track of it

---

[93] Chalmers 1996, p. 138.
[94] Chalmers 1996, p. 138.

all, but inconvenience is also consistent with truth. So, once again we get an objection from Chalmers that is not as good as his first objection. That's okay, since his first objection is a fairly compelling reason to reject MD.

The issue of whether MD is true is not settled. For one thing, Chalmers' arguments against MD are not obviously conclusive, and there are a number of defenses of MD that we haven't discussed. Before turning to defenses of MD, however, I would like to correct an important misconception. The misconception is to think that defending Chalmers' CP principles is as simple as defending ¬MD.

Is defending Chalmers' CP principles as simple as defending ¬MD?

We saw in the previous section that Chalmers thinks that defending his CP principles is as simple as defending ¬MD. For he thinks that his CP principles can be false only if there are strong necessities, and that there can be strong necessities only if MD is true. Accordingly, if MD is false, then there are no strong necessities, in which case there are no counterexamples to his CP principles, in which case his CP principles are true.

But do the sort of strong necessities that would be counterexamples to Chalmers' CP principles really require MD? I don't think so. I argue instead that, while rejecting MD is important for other purposes, MD's falsity doesn't guarantee the truth of his CP principles. I will explain this next.

Rejecting MD is an important part of Chalmers' argument, as it is required in order to justify his two-dimensional semantics. After all, if we were not persuaded of MD's falsity, we should be suspicious of a semantics whose intensions are written over the space of logically possible worlds. Rather, in so far as we take MD seriously, we should prefer a semantics where the intensions are written over a more restricted space of metaphysically possible worlds.

But, although rejecting MD is an important part of Chalmers' two-dimensional argument against materialism, ¬MD does not appear to imply Chalmers' CP principles. For, as we have seen, Chalmers' CP principles claim that ideal 1-conceivability entails 1-possibility, and thus a counterexample would be a statement that is ideally 1-conceivable yet 1-impossible. A counterexample would therefore imply a strong necessity. For, where S is a counterexample to Chalmers' CP principles, ¬S will be a strong necessity, as $\neg S_1$ will be necessarily true yet S_1 (its negation) will be ideally conceivable. Thus ¬CP iff there is a strong necessity in the following sense:

SN   For some S, $\Box S_1$ but $\neg S_1$ is ideally conceivable.

Since ¬CP≡SN, CP≡¬SN. Yet, although Chalmers defends ¬MD, it does not appear that he has *ipso facto* defended ¬SN. For the only way Chalmers has *ipso facto* defended ¬SN is if SN implies MD. For, if SN implies MD, then ¬MD (which he defends) implies ¬SN.

Does SN imply MD? To investigate this, consider the three collections at issue: (A) statements that are ideally conceivable when evaluated by 1-intensions; (B) statements that are possibly true when evaluated by 1-intensions; and (C) statements that are possibly true when evaluated by 1′-intensions, where 1′-intensions are just like 1-intensions, except that they are written over the space of metaphysically possible worlds. SN says that some statements are in collection A but not in B, while MD says that some statements are in B but not in C. Accordingly, SN⊃MD implies the following: if some statements are in collection A but not in B (SN's claim), then some statements are in collection B but not in C (MD's claim). But it is difficult to see why this might be true. Why might being in collection A but not in B imply that some statements are in B but not in C? There are only two options.

First, there is the option that any statement that is in collection A but not in B is *itself* in B but not in C. But this is impossible, as such a statement would be both in and not in B.

Second, there is the alternative option that, for any statement that is in collection A but not in B, a *different* statement is in collection B but not in C. Yet it is mysterious why a statement of the first sort should imply a different statement of the second sort. Why should S_1's being ideally conceivable but not possible imply that some other statement S′ is such that $\Diamond S'_1 \& \neg \Diamond S'_1$? We've already seen that the latter kind of statement would require the space of metaphysical possibility to have arbitrary restrictions built in, which is a feature that Chalmers criticizes in his critique of MD. So, it is clear that a statement that is in collection B but not in C would imply this kind of metaphysical arbitrariness. But why would a statement's being in collection A but not B imply that some different statement is in B but not in C, thereby requiring this metaphysical arbitrariness? Particularly, given that being in A but not in B does not involve metaphysical arbitrariness, why should it imply that some other statement involves metaphysical arbitrariness? The only option that comes to mind is that perhaps being in A but not in B is impossible and therefore the claim that some statements are in A but not in B (i.e., SN's claim) is false and thus vacuously implies everything, including implying that some different statements are in B but not in C. Yet this is an option according to which being in A but not in B (i.e., SN's claim) is impossible, which implies that SN is false, which implies that CP is true. So, on this option, CP is true regardless of whether MD is true, in which case defending ¬MD doesn't help defend Chalmers' CP principles.

In sum, there are two options for how SN might imply MD. The first option is impossible; meanwhile, on the alternative option, defending ¬MD doesn't help defend CP. Either way, defending ¬MD does not help defend ¬SN.

Perhaps a more intuitive way to state the problem is to note that collections A and B appear to be kinds of logical possibility. Collection B is called the space of logical possibility, while collection A can be thought of as a space of logical possibility by virtue of qualifying the conceivability as ideal, which should take the conception's quality to the limit of a kind of logical perfection. Yet SN requires that some statements be in collection A but not in B, in which case A and B are two distinct spaces of logical possibility. This appears to require a kind of logical arbitrariness. For some statements are logically possible in the one sense while being logically impossible in the other sense, but there don't seem to be any good objective grounds for this difference. Why should the logical perfection of ideal conceivability allow for more possibilities than are allowed by what we ordinarily call the space of logical possibility? Again, it is difficult to see why there should be this difference. Thus for a statement to meet SN's requirement of being in collection A but not in B to imply that some other statement meets MD's requirement of being in collection B but not in C, this would require SN's logical arbitrariness to imply MD's metaphysical arbitrariness. Yet it is difficult to see why this might be the case, unless it's because SN's logical arbitrariness is impossible and therefore SN is false and thus vacuously implies everything including MD. But, if SN is false, then there are no counterexamples to CP, in which case Chalmers' CP principles are true regardless of whether MD is true. In this case, defending ¬MD doesn't help defend CP, since CP is true regardless of whether MD is true.

Altogether, then, Chalmers is right that SN is the real threat to his CP principles, yet he is wrong to think that ¬MD implies ¬SN. Thus his strategy of defending ¬MD as a way of defending ¬SN is misguided. Instead of defending ¬MD, what he really needs to do is to give a direct critique of SN, perhaps explaining why SN seems false or incoherent.

Let us keep this in mind as we discuss critiques of Chalmers' CP principles. Chalmers discusses many such critiques,[95] but I would like to discuss some of them in more detail. As we will see, Chalmers is not the only philosopher who believes that SN implies MD.

Mathematical strong necessities?

There is an important objection regarding certain mathematical truths, which Chalmers says is "the most significant challenge."[96] Mathematical truths are generally taken to be logically necessary. Yet for all we know the negations of some of these truths might be ideally negatively conceivable. For instance, Goldbach's conjecture (or its negation) might be this way, or the continuum hypothesis (or its negation) might be this way. But if the negations of some necessary truths are ideally negatively conceivable, then CP- is false. For in that case ideal negative conceivability does not entail possibility.

Chalmers is skeptical of mathematical strong necessities, and he replies as follows: "In any given case, one can argue that either the statements in question are knowable under some idealization of rational reasoning, or that the statements are not determinately true or false."[97] Thus his response amounts to saying that it is plausible that every mathematical claim that *has* a determinate truth value can be known via *ideal* reasoning.

Whether Chalmers is right might depend on the correct theory of mathematics. According to Brouwer's intuitionism, a given mathematical claim is true only if it is knowable. Thus every mathematical truth is knowable, and thus it cannot be *ideally rational* to conceive the negation of a mathematical truth. Thus, if Brouwer's intuitionism is right, no mathematical truth can be a counterexample to CP-.

---

[95] See Chalmers 2010, p. 170-80.
[96] Chalmers 2002a, § 8 (ii).
[97] Chalmers 2002a, § 8 (ii). See also Chalmers 2010 p. 174-5.

By contrast with Brouwer's intuitionism, classical philosophy of mathematics holds that mathematical truths are necessarily true regardless of whether they are knowable. Additionally, we should be modest and admit that (for all we know) some mathematical truths might be knowable only by an infinitely long proof. For instance, a proof of Goldbach's conjecture might require a brute force search of all of the natural numbers, which, to use James F. Thomson's (1954) terminology, would be a *supertask*: i.e., it would require an infinite number of operations to be performed in a finite time. Although some philosophers argue that it is at least logically possible to complete a supertask,[98] others argue that completing a supertask is impossible.[99] This is a difficult issue, and there is no room to debate it here. Let us simply note that, if it is impossible to complete a supertask, this might prevent some mathematical truths from being knowable. Their unknowability would imply that their negations cannot be ruled out a priori and thus are ideally negatively conceivable, which would render them counterexamples to CP-. So, if some mathematical truths are unknowable—e.g., because completing supertasks is impossible— then there are strong necessities that render CP- false.

Altogether, if (a) classical philosophy of mathematics is true, and (b) some mathematical truths can be proven only by completing supertasks, but (c) supertasks cannot be completed, then it follows that there are mathematical strong necessities that render CP- false. There doesn't seem to be any compelling way of defending CP- in light of this objection. Rather, it seems that the best that can be said in defense of CP- is that, for all we know, (a), (b), and (c) are not all true, and that there are no other reasons to think that some mathematical truths are unknowable.

It is important to note, however, that this objection is only supposed to threaten CP-. By contrast, CP+, which yields a more plausible inference to possibility anyway, is supposed to

---

[98] E.g., Benacerraf 1962; Earman and Norton 1993, 1996.
[99] E.g., Thomson 1954; Black 1950-1; Burke 2000.

remain unscathed. Chalmers doesn't explain why this is: he simply says, "there is no challenge to CP+ here;"[100] but presumably the reason is as follows. In the case of a mathematical conception, a positive conception consists of a conceived proof. But a proof about mathematics cannot yield a false conclusion, as a proof can yield a false conclusion only if there is a mistake in the reasoning (a cognitive defect at play), which would require non-ideal reasoning.

There is more that can be said about the mathematical truth objection. A full treatment would delve deeply into specific potential counterexamples to CP-; but enough has been said to see that whether or not there are any mathematical strong necessities is a difficult issue, and there are no clear counterexamples to CP- here.

Strong necessities clouded by epistemic vagueness?

Timothy Williamson (1994) argues that there are necessities that are unknowable by virtue of being clouded in epistemic vagueness; although he does not apply his thesis to the issue of whether Chalmers' CP principles are true, it can be so applied. For, if the negations of Williamson's unknowable necessary truths are ideally conceivable, the unknowable necessities will be strong necessities.

To see this in detail, note that Williamson defends an "epistemic view of vagueness" according to which all vagueness is merely epistemic. On his view, the truth values of vague statements (e.g., "Tim is bald" or "that is a heap") are determinately true or false, yet unknowable. Metaphysically, there are sharp boundary lines between the bald and the non-bald, the heaps and the non-heaps, etc.; yet it is impossible to discover these sharp boundary lines. Moreover, these sharp boundary lines are supposed to be necessary. For Williamson accepts the supervenience thesis that "vague facts supervene on precise ones,"[101] and the supervenience is

---

[100] Chalmers 2010, p. 175.
[101] Williamson 1994, p. 202.

such that, "If two possible situations are identical in all precisely specified respects, then they are identical in all vaguely specified respects too."[102] He gives the following example:

(*)   If x has exactly the same physical measurements in a possible situation s as y has in a possible situation t, then x is thin in s if and only if y is thin in t.[103]

This leads to necessities. For, as Williamson puts it: "Let my exact physical measurements be m. According to the epistemic view, I am either thin or not thin. By (*), if I am thin then necessarily everyone with physical measurements m is thin."[104] Thus suppose I have measurements m and am thin. Then this, combined with (*), implies the following necessity:

MT   Everyone with physical measurements m is thin.

MT is necessarily true. Yet, on Williamson's epistemic theory of vagueness, MT might also be unknowable, given the unknowability of the boundary lines for what counts as thin and what doesn't. Moreover, assuming MT is unknowable, ¬MT might be ideally negatively conceivable, since this merely requires ideally conceiving it while failing to detect a contradiction in ¬MT. At any rate, if it *were* possible to detect a contradiction in ¬MT, MT should be knowable after all. All a person would have to do is detect a contradiction in ¬MT, and deduce MT. Thus ¬MT might be ideally negatively conceivable; yet, since MT is necessarily true, ¬MT is impossible. So, ¬MT would be ideally negatively conceivable, yet impossible. Thus ¬MT would be a counterexample to CP-, and MT would be a strong necessity.

Yet I argue that this does not yield a counterexample to CP-. For a counterexample to CP- would be a statement that, when evaluated by 1-intensions, is both ideally negatively conceivable and impossible. Thus, for ¬MT to be a counterexample to CP-, ¬MT would have to be both ideally negatively conceivable and impossible when evaluated by a 1-intension. That is,

[102] Williamson 1994, p. 202-3.
[103] Williamson 1994, p. 202-3.
[104] Williamson 1994, p. 203.

$\neg MT_1$ would have to be both ideally negatively conceivable and impossible, in which case MT_1 would have to be necessarily true.

However, I think it is absurd to claim that MT_1 is necessarily true. For I think this would require a kind of arbitrariness that is impossible. Particularly, necessary truths clouded by epistemic vagueness (if there are any such truths) seem logically arbitrary. For some sharp boundaries clouded by epistemic vagueness might be, say, here or there, yet logically it seems that such boundaries could just as well have been elsewhere. Yet the necessary truth of MT_1 would have to be a result of logically arbitrary boundary lines. After all, for a statement S to be necessarily true when evaluated by a 1-intension is for the following to be the case: there must be *no* logically possible world where S_1 is false. Yet, if the relevant boundary lines really are logically arbitrary, then there *should* be a logically possible world where S_1 is false. For saying that something is logically arbitrary is precisely to say that there are logically possible worlds where it holds, and logically possible worlds where it doesn't hold.[105] Thus, given the logical arbitrariness of the sharp boundary lines that ground the necessities that are clouded in epistemic vagueness, $\neg MT_1$ should not be impossible, in which case $\neg MT_1$ should not be both ideally negatively conceivable and impossible, thereby contradicting CP-. Generalizing on $\neg MT_1$, I conclude that it is impossible for there to be a statement, S_1, that is ideally negatively conceivable yet impossible, whereby $\neg S_1$'s necessity is unknowable due to being clouded in epistemic vagueness. For any such statement will absurdly require that a strong necessity's necessity be logically arbitrary, hence logically contingent, hence not logically necessary. Any such necessity's necessity, then, will be contradicted by its own possible falsity. Thus it is

---

[105] This is a bit charitable. For, strictly, a statement's being logically arbitrary might also be taken to mean that it is neither determinately true nor determinately false when evaluated by intensions that are written over the space of logically possible worlds. Yet, on this option, necessities clouded by epistemic vagueness could not help ground counterexamples to CP- anyway, as counterexamples to CP- would have to be determinately impossible.

impossible for there to be a strong necessity whose necessity is unknowable by virtue of being

clouded in epistemic vagueness. Hence no epistemically vague statement can serve as a

counterexample to Chalmers' CP principles.

Here ends my critique of the claim that the epistemic theory of vagueness might be a

source of counterexamples to CP-. Chalmers gives a different critique. Particularly, he (e.g.,

Chalmers 2002a, 2005, 2010) mentions the epistemic theory of vagueness as a potential source

of counterexamples to CP-, but he points out that "all this is contingent on the truth of the

epistemic theory of vagueness, and the epistemic theory is widely regarded as very

implausible."[106] He then gives a specific critique of the epistemic theory of vagueness, by appeal

to the example of *baldness*. He writes:

> One might trace the implausibility of the epistemic theory at least in part to [the fact that]
> it seems that a subject has all the qualitative information that could possibly be relevant,
> and it seems almost obvious that given that information, the subject is in a position to
> know all there is to know about baldness here.[107]

The phrase "almost obvious" is key. Why should we agree with Chalmers that it seems, "almost

obvious that given that information, the subject is in a position to know all there is to know about

baldness here"? Chalmers doesn't elaborate on why it should seem almost obvious. Yet I think it

can be articulated. In fact, we might want to upgrade it to "downright obvious."

For starters, Chalmers and Williamson agree that heads of hair are non-vague: they are

precise. Thus to know the precise details of a head of hair is to know all there is to know about it

per se. If we try to learn all about it, we might start concerning ourselves with relations. For

instance, we might take an interest in all the possible ways that a given head of hair can be

arbitrarily categorized at macro levels: here we might be interested in all the logically possible

ways of defining a heap of hair on a head, or of defining what counts as a greasy head of hair, or

---

[106] Chalmers 2002a, § 8 (ii).
[107] Chalmers 2002a, § 8 (ii).

of defining what counts as a bald head. But here we would have numerous arbitrary options for how to make the distinctions. Moreover, the possibility of such categorizations is easily knowable and any one of them can be stipulated for any reason. Finally, we might be interested in the relations between heads of hair and epistemically vague statements about heads of hair. Yet the crucial point here is that, no matter what precise distinctions the epistemically vague predicates might make with regards to heads of hair, they will be distinctions that we could just as well have arbitrarily stipulated from the outset. Thus they cannot be categorizations that shed any light on what we already were in a position to see as an arbitrary categorization of what we already knew was there. Thus the unknowable necessity, if there is one, cannot be an unknowable necessity about heads of hair *per se*; it must rather be an unknowable necessity about epistemically vague *statements* about heads of hair. Particularly, it must be an unknowable necessity about which *arbitrary* classifications those vague statements pick out. This renders it not a fact about baldness *per se*, but merely a fact about *reference*, given the mysterious ways in which epistemically vague utterances refer.

Williamson might nevertheless reply to Chalmers that, although it is not a fact about baldness, and instead is a fact about reference, it is nevertheless a fact! Yet even this can be plausibly disputed. For the view that epistemically vague kind terms pick out non-vague sets is highly counterintuitive. That is, the following principle is highly counterintuitive:

Although vague predicates do not specify precise membership conditions for the sets they are talking about, nevertheless they pick out sets with precise membership conditions.

This is problematic. For it is nearly irresistible to think that a vague predicate is the wrong kind

of tool with which to pick out a non-vague referent. It seems analogous to using a cookie cutter that lacks a definite shape as the tool with which to carve out chunks of dough that all have exactly the same perimeter. Either way—whether it's a vague predicate picking out a non-vague set, or an indistinctly-shaped cookie cutter carving out a distinct perimeter—the process by which the imprecise picks out the precise is mysterious. We might call this *The Vague Cookie Cutter Problem*. Likewise, we might call the thesis that is indented above for emphasis, and which generates *The Vague Cookie Cutter Problem*, *The Vague Cookie Cutter Thesis*.

Although *The Vague Cookie Cutter Problem* is serious, *The Vague Cookie Cutter Thesis* should not be dismissed out of hand; after all, Kripke's well received theory of reference appears to have the same sort of problem. For Kripke's theory requires some sort of "reference magnetism" so that a given name or gesture, however indistinct it may be, manages to pick out (say) a natural kind, as opposed to some arbitrary non-natural pattern of particles; or that it manages to pick out (say) Jones, without also picking out the neutrinos that are currently running through Jones's body. Given how well received Kripke's theory is, maybe we should take Kripke's theory as fact, and use it to run a reductio argument against the seriousness of the vague cookie cutter problem. But I think this would be premature; for maybe Kripke's theory is wrong. After all, let's be honest, how is the activity of naming or (say) finger pointing supposed to make all of these "choices" about what to consider relevant to the reference, or what to consider as essential or accidental to the reference? Calling it "reference magnetism" is, from a *conceptual* standpoint, to give no account at all. In my view, this is one of the reasons Williamson's epistemic view of vagueness is so implausible: namely, it proposes that the referent picked out is the result of what seems to require *decisions* that were not made as part of the linguistic activity of picking it out, as the linguistic activity of picking it out was (by hypothesis) imprecise. Thus

the epistemic theory of vagueness seems to imply that the decisions are made, but that the process by which they are made is a mystery. Kripke's theory of reference is the same way. So, although it is tempting to throw Williamson's epistemic theory of vagueness under the rug, it is difficult to see how this could be done with consistency without throwing Kripke's theory of reference under the rug, which most philosophers are reluctant to do. Yet the problem is serious, and maybe a talented defender of Kripke's theory of reference can show us how Williamson can defeat *The Vague Cookie Cutter Problem*.

Altogether, it seems that, if there are necessities that are unknowable due to epistemic vagueness, they cannot imply counterexamples to Chalmers' CP principles. Moreover, there are two reasons to think that there are no such necessities. First, I argued that any such necessity (if there are such necessities) would be impossible, as it would require that a certain kind of logical arbitrariness be both present and absent. Second, Chalmers argues that, beyond the precise facts about the world, there doesn't seem to be anything else to know about it. I argued that Chalmers' argument can fail only if *The Vague Cookie Cutter Thesis* is true, which is implausible, since that thesis has *The Vague Cookie Cutter Problem*. Finally, in addition to the fact that the epistemic theory of vagueness is plausibly both irrelevant and false, it is important to keep in mind that it is at best only a threat to CP- anyway.

Strongly necessarily existing God?

Another objection to Chalmers' CP principles regards the possibility of a necessarily existing God. Particularly, one might argue that both the existence and non-existence of a necessarily existing God is ideally conceivable; so (by CP) such a God is both possible and impossible. By reductio, CP is false.[108]

---

[108] Stephen Yablo gives a similar argument in his (1999) paper, "Concepts and Consciousness," although he doesn't qualify the conceivability as *ideal* conceivability.

This argument is deeply implausible. For it is deeply implausible that a necessarily existing God is ideally conceivable. Indeed, this is Chalmers' response. For starters, such a God doesn't seem positively conceivable: as Chalmers puts it, "I can certainly form no clear and distinct conception of such a god."[109] This is understandable: for even theists typically admit that they have no idea how to positively conceive of a God, let alone a necessarily existing God. Thus it seems that, at best, such a God might be ideally negatively conceivable.

But a necessarily existing God doesn't seem ideally negatively conceivable either. For one thing, it's not obvious that the concept of a God is coherent even as a being that exists in some worlds, let alone all worlds. For even in a single world, there are serious questions about how a God might know anything, or do anything, without having a body or a brain, and perhaps without even being in time. Chalmers is presumably thinking along these lines when he says, "rational reflection reveals all sorts of problems with the idea."[110] These problems are evidence that a God is not even ideally negatively conceivable, as it is evidence that an ideal reasoner, or someone reasoning in an ideal way, would consider the concept of a God incoherent.

To make matters worse, an atheistic world seems easily conceivable. Indeed, billions of atheists have an easy time conceiving of our world in an atheistic way. This is evidence that an atheistic world is ideally conceivable, which would preclude the ideal conceivability of a necessarily existing God. So, it is plausible that a necessarily existing God is not ideally (positively or negatively) conceivable, and thus that there is no threat to CP here.

Vaidya's "essential origins" argument against CP+

Let us now turn to Anand Vaidya's (2008) argument against CP+. Vaidya defends MD, and from here he argues that CP+ is false. His defense of MD requires first arguing that, metaphysically,

---

[109] Chalmers 2010, p. 179.
[110] Chalmers 2010, p. 179.

¬(for any x, □x⊃□□x). That is, he argues against the characteristic axiom of S4 logic.[111] We can abbreviate this by saying that he defends ¬S4.[112] He then uses ¬S4 in a defense of modal dualism (MD), which he uses in a defense of ¬CP+. I will now discuss Vaidya's arguments in this order: first his defense of ¬S4; then his defense of MD; and finally his defense of ¬CP+.

S4 is the thesis that, metaphysically, for any x, □x⊃□□x. For short, S4 is the view that necessity implies necessary necessity. It is a popular principle. Yet Vaidya rejects this principle. His argument against it cites the principle of the essentiality of origins; for short, *Origin Essentialism*. On Vaidya's terminology, *Origin Essentialism* is the view that everything has its origin essentially, so that it is not possible for anything to have a different origin than the origin it in fact had. Vaidya only defines it once: he says, *"the essentiality of origin—the principle that the origin of x is a property x has in every possible world in which x exists."*[113]

Origin Essentialism, which Vaidya accepts, is an alternative to an extreme kind of haecceitism that some philosophers find problematic. For instance, David Lewis raises the issue of whether a person could have been a poached egg (thus presumably originating from a hen). "The most extreme version [of haecceitism]," he says, is "that . . . anything could have any qualitative character, for instance there is a . . . [possible] world according to which you are a poached egg."[114] Yet many philosophers argue, or intuit, that no human could have been an egg. For instance, L. A. Paul writes, "I am not possibly a fried egg." Chicken eggs aside, Kripke holds that no human could have come from a different sperm and egg than he or she in fact came from, and therefore that Queen Elizabeth could not have come from a "*totally* different sperm and

---

[111] So dubbed by C. I. Lewis.
[112] This is the abbreviation that Vaidya uses, although it is potentially misleading, as he doesn't think S4 logic is false of logical modality; he only thinks it's false of metaphysical modality.
[113] Vaidya 2008, p. 200.
[114] Lewis 1986, p. 239

egg."[115] Kripke also holds that the wooden table in the room could not have come from a "completely *different* block of wood, or . . . of water cleverly hardened into ice-water taken from the Thames river."[116] Thus some philosophers find themselves in this general tradition of being opposed to the extreme haecceitism that Lewis discusses. Indeed, Lewis himself is opposed to this extreme haecceitism, as is Paul, Kripke, Vaidya, and many others. One alternative to this extreme haecceitism is *Origin Essentialism*. Yet, even if one rejects the possibility of extreme origin variation that is implied by this extreme haecceitism, one might still accept the possibility of minor origin variation.

Indeed, despite accepting *Origin Essentialism*, Vaidya allows for slight origin variation, which we might call *Origin Tolerance*. Thus Vaidya accepts both *Origin Essentialism* and *Origin Tolerance*. This might seem to be a tense combination. Indeed, I will argue that *Origin Essentialism* and *Origin Tolerance* are mutually inconsistent. First, however, I will explain Vaidya's defense of ¬S4. In this exposition, I will explain his notion of *Origin Tolerance*, and I will explain why he takes *Origin Essentialism* and *Origin Tolerance* to be consistent.

Vaidya's motivation for ¬S4 is to avoid a contradiction. For suppose,

0. | "There is a tree T from which a portion of wood m is carved and completely
      | used to make a table t."[117]

This hypothesis, combined with the next three lines, entails a contradiction:

1. | For any x, $\Box x \supset \Box\Box x$ [S4]

2. | $\Box$(if t exists, t came from m)   [0 *Origin Essentialism*]

3. | $\Diamond\Diamond$(t exists, but t came from m*, where m ≠ m*)   [0 *Origin Tolerance*]

A contradiction can be derived as follows:

---
[115] Kripke 1972/2001, p. 113.
[116] Kripke 1972/2001, p. 113-4.
[117] Vaidya 2008, p. 200.

4. | □□(if t exists, t came from m)   [1, 2]

5. | ¬□¬◊(t exists, but t came from m*, where m ≠ m*)   [3 modal negation]

6. | ¬□□¬(t exists, but t came from m*, where m ≠ m*)   [5 modal negation]

7. | ¬□□(¬ t exists, or ¬ t came from m*, where m ≠ m*) [6 DeMorgan]

8. | ¬□□(if t exists, ¬ t came from m*, where m ≠ m*)   [7 conditional exchange]

9. | ¬□□(if t exists, ¬ t didn't come from m)   [8 analytic]

10. | ¬□□(if t exists, t came from m)   [9 double negation]

Line 10 negates line 4.

One can avoid this contradiction by denying any of lines 1-3. Most philosophers accept line 1; so most philosophers would reject line 2 or 3. Nevertheless Vaidya argues that the best way to avoid the contradiction is to deny line 1, as he believes that our intuitions in favor of lines 2 and 3 ought to be much stronger than our intuitions in favor of line 1. Why does Vaidya think lines 2 and 3 are worth preserving? Unfortunately, the answer can only be given partly, as he never defends line 2: he takes *Origin Essentialism* as a given. Nevertheless, he does explain why he thinks we should hold on to line 3, which is a consequence of his *Origin Tolerance*. Namely, he says that accepting lines 1-2 while rejecting line 3 would have the following problem:[118]

> The main problem [is] that in maintaining the essentiality of origin while denying that it is possibly possible that t originate from m* one is committed to maintaining that every bit of matter that was actually at the origin is relevant to individuating t from some other table t*, and essential to t's identity. It is this claim that appears so hard to accept.[119]

The "main problem," which is the only problem he mentions, is that accepting lines 1-2 while rejecting line 3 commits one to the following consequence:

---

[118] I am giving Vaidya the benefit of the doubt. For he says that the problem results if one accepts line 2 while rejecting line 3; yet, strictly, if modal accessibility is restricted (and he thinks it is, which we'll discuss shortly), the problem might require accepting both lines 1 and 2 while rejecting line 3.
[119] Vaidya 2008, p. 201.

>*All Matter Matters*: "every bit of matter that was actually at the origin is relevant to individuating t from some other table t*, and essential to t's identity."

Vaidya is loathe to accept *All Matter Matters*, as he argues that *Origin Essentialism* should allow for some "vagueness" as to which matter is relevant to table individuation and identity. For he argues that "[m]etaphysical modality is a kind of modality that flows from the nature of the entities in question,"[120] but that "[w]hat we [rightly] care about at the origin when individuating an object changes with the kind of object we are dealing with,"[121] and that, when the object in question is an artifact such as a table, some "vagueness . . . is tolerable at the origin."[122] In this way, Vaidya defends *Origin Tolerance*, which is inconsistent with *All Matter Matters*. Thus, if accepting lines 0 and 2 while rejecting line 3 implies *All Matter Matters*, then so much the worse for accepting lines 0 and 2 while rejecting line 3.

Vaidya argues that the best option is to reject line 1. After all, lines 0, 2, and 3 are obviously true. Yet lines 1 and 2 together imply *All Matter Matters*,[123] which is inconsistent with *Origin Tolerance*, in which case it is also inconsistent with line 3. (Absurd.) By reductio, lines 1 and 2 are not both true. But line 2 is obviously true. So, line 1 is false. That is, ¬S4.

I argue that Vaidya's defense of ¬S4 is unsound on the grounds that lines 2 and 3 are inconsistent. Lines 2 and 3 are applications of *Origin Essentialism* and *Origin Tolerance*, and they are inconsistent if one says that t has origin m in every world where t exists, while the other says that t does not have origin m in every world where t exists. For in that case, they would imply that t both does and doesn't have origin m in every world where t exists. However, Vaidya

---

[120] Vaidya 2008, p. 202.
[121] Vaidya 2008, p. 202.
[122] Vaidya 2008, p. 202.
[123] Proof: line 1's claim, S4, can be applied to line 4's claim, □(if t exists, t came from m), to derive, □□(if t exists, t came from m), and again to derive, □□□(if t exists, t came from m), etc., ad infinitum, thus *All Matter Matters*.

tries to avoid the inconsistency by accepting "restrictions on accessibility" between possible

worlds; for short, he accepts that modal accessibility is restricted. David Lewis explains it well:

> All hands agree that very often our modalities are quantifications restricted to 'accessible' worlds - we tacitly ignore worlds where the past differs, where the actual laws of nature are violated, where there are alien natural properties, or what have you. . . . We [ignore] far-out worlds where things differ too much in qualitative character from the way they actually are. Almost always, apart from philosophical discussions that can scarcely be put in evidence as samples of 'linguistic intuition', at least these far-out worlds are left aside as inaccessible. Leaving them aside, indeed you could not have been a poached egg.[124]

To say that modal accessibility is restricted, then, is to say that what is possible is a relative

matter, where possibility is restricted to the worlds that are "nearby" in the relevant sense. Thus a

given statement is possibly true iff it is true in at least one nearby world, regardless of whether it

is true in "far-out" worlds. The consistency of lines 2 and 3, then, is supposed to consist in the

following. *Origin Essentialism* is only a claim about what is metaphysically possible in the

nearby worlds, while *Origin Tolerance* is a claim about a larger class of worlds. Thus *Origin

Essentialism* and *Origin Tolerance* are consistent; for *Origin Tolerance* implies that table t exists

without having origin m in some of the worlds in this larger class, whereas *Origin Essentialism*

merely says that t has origin m in all of the nearby worlds.

Vaidya notates this by letting one iteration of "necessarily" or "possibly" denote the

nearby worlds, while adding a second iteration (e.g., "possibly possibly") to expand the modal

reach beyond the first restricted class so as to include any world that is nearby any of the nearby

worlds. Thus if we let "X" stand for the nearby worlds, and let "Y" stand for the worlds that are

nearby some of the nearby worlds, X would be a proper subset of Y. As such, *Origin

Essentialism* and *Origin Tolerance* are consistent; for *Origin Essentialism* says that table t has

---

[124] Lewis 1986, p. 240.

origin m in every X-world where t exists; yet, even if this is true, *Origin Tolerance* might still be

right that t lacks origin m in at least one Y-world where t exists, as Y is a proper superset of X.[125]

So, the consistency of lines 2 and 3 requires that modal accessibility be restricted. Yet I

argue that modal accessibility is unrestricted. One problem with holding that modal accessibility

is restricted is that it implies that we are condemned to *ignore* certain worlds that are included in

the most all-encompassing collection of worlds. But it's not clear why we must ignore them. If

they are included in the most all-encompassing collection, why can't we quantify over them as

such? As David Lewis puts it, "by what right do we ignore worlds that are deemed inaccessible?

Accessible or not, they're still worlds. We believe in them. Why don't they count?"[126] The point

here is that it seems arbitrary to say that we can't quantify over all of them. Moreover, there

appears to be a further arbitrariness problem as well. Namely, it seems to me that any *metric* by

which worlds are laid out will also be arbitrary, as some worlds that are deemed accessible in one

metric will be deemed inaccessible in a different, equally legitimate metric. Thus accepting

restrictions on modal accessibility has two arbitrariness problems, which should constitute a

reductio against restrictions on modal accessibility. Thus I conclude that modal accessibility is

unrestricted, in which case lines 2 and 3 are inconsistent.

---

[125] It is interesting that Vaidya's *Origin Tolerance* appears to only extend so far. After all, class Y might not be the most all-encompassing class of worlds. If Y is not the most all-encompassing class of metaphysically possible worlds, then a third iteration—e.g., "possibly possibly possibly"—would bring us even further into the depths of metaphysical possibility space. It would bring us to any world that is nearby any of the worlds in Y. We could call this new class of worlds "Z." And so it goes: adding more iterations of "possibly" or "necessarily" gradually brings us closer and closer to the most all-encompassing class of worlds. Why does Vaidya's *Origin Tolerance* extend only to the Y-worlds, and not necessarily to the Z-worlds, or to any world no matter how far out? That is, why does he only use two diamonds, and not three or more? Why does he not use as many diamonds as it takes to reach into the furthest depths of metaphysical possibility space, so as to say that there is at least one world, no matter how far out, where t exists but didn't come from m? This is unclear, although it won't have to concern us here. For, even if Vaidya were to extend it further, the modified formulation of *Origin Tolerance* would still be inconsistent with the conjunction of lines 1 and 2. For, however many diamonds the modified formulation of *Origin Tolerance* incorporates (say, *n*) it will be equivalent to the negation of *n* boxes in front of the claim that, if t exists, t came from m. This will negate a formula that can be derived from lines 1 and 2. Particularly, if *Origin Tolerance* incorporates *n* diamonds, this will be equivalent to the negation of a statement that can be derived from lines 1 and 2 by *n*-1 appeals to S4 (line 1).
[126] Lewis 1986, p. 246.

Vaidya would reply to these arbitrariness problems by arguing that lines 2 and 3 are true, and therefore mutually consistent. His justification for line 2 is the appeal to *Origin Essentialism*, while his justification for line 3 is the appeal to *Origin Tolerance*. Unfortunately, however, he doesn't defend *Origin Essentialism*, and his defense of *Origin Tolerance* is an appeal to intuition. He writes:

> Artifacts, such as tables and statues, have their origin essentially, but they are such that it is possibly possible that they come from slightly different matter because concerning these kinds of objects we do not care about all the matter. This intuition about artifacts is a data point an adequate theory of essence must capture, and which itself constrains epistemologies of essence.[127]

Here there is no defense of *Origin Essentialism*, and the only defense of *Origin Tolerance* is that "concerning these kinds of objects we do not care about all the matter." On the face of it, this may make it seem as though Vaidya thinks the modal properties of artifacts are interest-dependent, influenced by "what we care about." In private correspondence, however, Vaidya rejects this interpretation. He does not mean to suggest that our interests influence their modal properties, but rather that their modal properties depend on their interest-independent natures. Perhaps the biggest clue in the text of this (correct) interpretation is that Vaidya refers to what we do not care about with respect to the origins of artifacts as "intuition." Being an intuition, the emphasis should be on its status as *aiming* at the truth, not as being something that might influence what is possible. Thus the idea is not that our intuition influences the modal properties of tables, but rather that our intuition apprehends truths upon which the modal properties of tables supervene. Indeed, Vaidya takes this intuition to be so compelling that he writes: "[t]his intuition about artifacts is a data point an adequate theory of essence must capture."

This is strong language. Ordinarily a data point is taken as unquestionably true. As such, Vaidya's defense of *Origin Tolerance* would be that, since we have such a strong intuition that

---

[127] Vaidya 2008, p. 203.

not all of the matter at the origin of artifacts matters, *Origin Tolerance* is unquestionably true. Thus Vaidya takes his intuition to be a reliable guide to possibility. Yet it is difficult to see why his appeal to intuition should be more plausible than CP+, especially as it has not been qualified as requiring the relevant intuitions to be ideal, positive, or primary. Thus his defense of *Origin Tolerance*, and thus line 3, appears to be less justified than the thesis that he aims to reject: CP+.

In sum, I argue that Vaidya's lines 2 and 3 are inconsistent, while his only potential reply seems to be to argue that lines 2 and 3 are true and therefore consistent. But his justification for line 2 is *Origin Essentialism*, which he doesn't defend, and his justification for line 3 is *Origin Tolerance*, which he defends only by appeal to intuition. Altogether, Vaidya has an uncompelling case against my argument that lines 2 and 3 are inconsistent. So, it remains plausible that Vaidya's defense of ¬S4 is unsound.

Recall that Vaidya's defense of ¬S4 is only the first part of his defense of ¬CP+. Let us now discuss the second part of his defense of CP+: his defense of MD. MD says that every logical possibility is metaphysically possible, but not every metaphysical possibility is logically possible. Vaidya holds that the only alternative to MD is the *Modal Monist* (MM) thesis that the space of metaphysically possible worlds is *identical* to (i.e., extensionally equivalent to) the space of logically possible worlds.[128, 129]

Vaidya defends MD as follows. Suppose for reductio that MM is true—that the space of metaphysically possible worlds is extensionally equivalent to the space of logically possible worlds. But any pair of extensionally equivalent sets must have the same characteristic axioms. So, since logical modality is characterized by the axiom S4, metaphysical modality must also be characterized by the axiom S4. But Vaidya just argued that metaphysical modality is not

---

[128] See Vaidya 2008, p. 196.
[129] Syntactically, there's a third permutation: namely, that the set of logically possible worlds is a proper subset of the set of metaphysically possible worlds. But nobody considers this option to be feasible.

characterized by the axiom S4. So, there is a contradiction: metaphysical modality is and isn't characterized by the axiom S4. By reductio, MM is false, in which case MD is true.[130]

Vaidya in turn employs MD in his defense of ¬CP+. Letting "SN" stand for the claim that there are some strong necessities, this defense can be outlined as follows:

(a)   MD⊃SN   ["Modal Dualism opens up Strong Necessities"[131]]

(b)   SN⊃¬CP+   ["Strong Necessities refute Modal Rationalism"[132, 133]]

(c)   CP+⊃¬MD   [(a), (b) hyp. syllogism, contraposition, double negation]

(d)   ¬MD⊃MM   ["Exclusivity of Relations"[134]]

(e)   CP+⊃MM   [(c), (d) hyp. syllogism]

(f)   ¬MM   [defended earlier]

(g)   ¬CP+   [(e), (f) modus tollens]

Lines (c), (e), and (g) are justified by truth-preserving inferences rules; the justification for line (d) was explained earlier; and line (f) is the conclusion of Vaidya's second argument: the one we just went over. Thus the only remaining work to do is to see how Vaidya justifies lines (a) and (b). I will argue that Vaidya's justifications for lines (a) and (b) are excessively weak.

Line (a) says that MD is true only if there is at least one strong necessity. For short, MD⊃SN. Vaidya justifies this by saying, "Modal Dualism opens up Strong Necessities." It is not clear what Vaidya means by "opens up," and he doesn't elaborate on it. Whatever it is supposed to mean, however, it is difficult to see how it might be true. I argued earlier in this

---

[130] See Vaidya 2008, p. 199.
[131] Vaidya 2008, p. 195.
[132] Vaidya 2008, p. 195.
[133] Vaidya 2008 uses "Modal Rationalism" synonymously with "CP+." For he defines it as follows: "(MR) Modal Rationalism: primary positive ideal conceivability entails primary possibility" (p. 192).
[134] Vaidya 2008, p. 195.

chapter that, if SN is true, SN doesn't imply MD; I will now argue that, if MD is true (as Vaidya maintains), then MD doesn't imply SN.

Does MD imply SN? To settle this question, consider the three collections at issue: (A) statements that are ideally conceivable when evaluated by 1-intensions; (B) statements that are possibly true when evaluated by 1-intensions; and (C) statements that are possibly true when evaluated by $1'$-intensions, where $1'$-intensions are just like 1-intensions, except that they are written over the space of metaphysically possible worlds. MD says that some statements are in B but not in C, while SN says that some statements are in A but not in B. Accordingly, MD⊃SN implies the following: if some statements are in collection B but not in C (MD's claim), then some statements are in collection A but not in B (SN's claim). But it is difficult to see why this might be true. Why might being in collection B but not in C imply that some statements are in A but not in B? There are only two options.

First, there is the option that any statement that is in collection B but not in C is *itself* in A but not in B. But this is impossible, as such a statement would be both in and not in B.

Second, there is the alternative option that, for any statement that is in collection B but not C, a *different* statement is in A but not B. Yet it is mysterious why a statement of the first sort should imply a different statement of the second sort. Why should it follow from the fact that, for some S, $\Diamond S_1 \& \neg \Diamond S_{1'}$, that some other statement S^*_1 is ideally conceivable but impossible? The former kind of statement requires that the space of metaphysical possibility has arbitrary restrictions built in, which Chalmers criticizes in his critique of MD. So, a statement that is in B but not C would imply this kind of metaphysical arbitrariness. Yet, as I pointed out earlier, being in A but not B seems to require a kind of logical arbitrariness. Thus, for a statement to meet MD's requirement of being in collection B but not C, and to imply that some other statement

meets SN's requirement of being in collection A but not B, this would require that MD's metaphysical arbitrariness implies SN's logical arbitrariness. Yet it is difficult to see why this might be the case, unless it's because MD's metaphysical arbitrariness is impossible and thus MD vacuously implies everything including SN. But if MD were impossible, MM would be true, which would contradict line (f), rendering Vaidya's argument unsound.

So, if MD is true, MD doesn't imply SN (i.e., line (a) is false); if MD is false, line (f) is false. Either way, Vaidya's defense of ¬CP+ is unsound.

I conclude that Vaidya's defense of CP+ is problematic in four ways. For it involves defending ¬S4, MD, and ¬CP+, but each of these arguments is problematic. First, his defense of ¬S4 requires that metaphysical modal accessibility be restricted, which seems arbitrary. Second, his defense of ¬S4 appeals to an intuition that seems less justified than the thesis that he aims to reject, CP+. Third, his defense of MD is vulnerable to Chalmers' critique of MD. Fourth, line (a) and line (f) cannot both be true, and thus Vaidya's defense of ¬CP+ is unsound.

Howell's critique of "metaphysical two-dimensionalism"

Robert Howell (2008) rejects Chalmers's CP principles on the grounds that they imply "metaphysical two-dimensionalism," which he defines as, "any view that combines two-dimensionalism with the view that there are always metaphysically possible worlds that make coherent primary intensions true.[135] He takes Chalmers' CP principles to imply metaphysical two-dimensionalism since, according to Chalmers' CP principles, "conceivability does deliver real possibilities."[136] Yet Howell rejects two-dimensionalism; for, as he puts it: "if metaphysical

---

[135] Howell 2008, p. 351.
[136] Howell 2008, p. 350.

two-dimensionalism is assumed, it can be shown to be false, and any move that might avoid this result makes two-dimensionalism of little dialectical use in metaphysical debates."[137]

Unlike many of Chalmers' critics, Howell is careful to argue against Chalmers' two-dimensional argument against materialism, as opposed to arguing against a more crude anti-materialist formulation that Chalmers has thrown out for initial consideration. Thus Howell explains Chalmers' two-dimensional semantics, and incorporates the primary/secondary intension distinction in his critique. Roughly, Howell's strategy is to argue that Chalmers' CP principles imply a contradiction. Recall Chalmers' CP principles:

CP+   Ideal primary positive conceivability entails primary possibility.

CP-   Ideal primary negative conceivability entails primary possibility.

Surprisingly, however, Howell does not distinguish between CP+ and CP-, as he does not even distinguish between positive and negative conceivability.[138] This is unfair to Chalmers and constitutes a distortion of Chalmers' argument.

A second distortion is that Howell neglects to qualify the relevant conceptions as *ideal*; for he argues that qualifying conceptions as ideal "is an invitation to a rather frustrating debate involving what amounts to little more than intuitional hearsay. One person can insist that they are conceiving of a possibility, only to be set straight by their opponent who claims that their conception is only prima facie."[139] Howell wants to eliminate the qualification of *ideal* from the formulation of the CP principles—apparently on the grounds that whether or not a given conception is ideal is not provable or knowable. Yet Howell completes his thought by saying, "Despite a sense that this dialectical dance [of arguing whether a given conception is ideal] is

---

[137] Howell 2008, p. 348.
[138] The word "positive" never appears in his paper, and the word "negative" appears only in a quoted passage where Chalmers uses the phrase "negatively conceivable" (but even here the qualification of *negatively* is then ignored in Howell's paper).
[139] Howell 2008, p. 353.

one that should be declined, I must say that it is not clear that we cannot conceive of the possibility . . . in a strong enough manner."[140] Strong enough for what? Presumably strong enough to entice Chalmers (or like-minded two-dimensionalists) into thinking that the given kind of conceivability (without the qualification of *ideal*) entails possibility.

Altogether, Howell believes that it is reasonable to ignore the *positive/negative* distinction and to eliminate the qualification of *ideal*, so that Chalmers' CP+ and CP- principles should be replaced with a different conceivability-to-possibility principle. He doesn't say how it should be spelled out, but I suppose it must go roughly as follows:

CP′   Coherent conceivability$_1$ entails possibility$_1$.

That is, $\Box$(for any statement, S, S$_1$ is coherently conceivable only if $\Diamond$S$_1$). Henceforth I will take CP′ as Howell's recommended alternative to the CP+/CP- distinction. That is, although he knows that Chalmers defends CP+ and CP-, he finds these principles untenable, and so he thinks that Chalmers should try defending CP′ instead. I am taking a liberty here, as Howell never gives this formulation, or any other. But it seems to be what he's after.

One of the key terms in Howell's argument is "SN," by which he means, "The space of metaphysically possible worlds is more limited than the space of conceivable worlds."[141] He writes: "I call this SN because it basically maintains that there are what Chalmers . . . calls 'strong necessities'."[142] Howell uses "SN" in a different way than I have been using "SN." I have been using "SN" in the following way,

SN   For some S, $\Box$S$_1$ but $\neg$S$_1$ is ideally conceivable.

whereas Howell uses "SN" in roughly the following way:

SN′   For some S, S$_1$ is coherently conceivable but not metaphysically possible.

---

[140] Howell 2008, p. 353.
[141] Howell 2008, p. 352.
[142] Howell 2008, p. 352, footnote 8.

Strictly, SN′ is probably misleading. For, although Howell uses the phrase "metaphysically possible worlds" and takes himself to be arguing against "metaphysical two-dimensionalism," he recognizes that Chalmers' two-dimensional semantics is supposed to operate over a single space of possible worlds, and presumably he realizes that this space of worlds is what philosophers ordinarily call the *logically possible* worlds. For charity's sake, then, and for extra clarity, I think it is helpful to reformulate Howell's "SN" thesis as follows:

SN″   For some S, S_1 is coherently conceivable but not logically possible.

That is, for some S, S_1 is coherently conceivable but $\neg \lozenge S_1$. If there is such an S, $\neg S_1$ will be necessarily true despite the fact that its negation S_1 is coherently conceivable. This does not obviously imply that there is a strong necessity, however, as its negation is not said to be *ideally* conceivable; it is merely said to be *coherently* conceivable. Thus Howell's claim that "[SN″] . . . basically maintains that there are what Chalmers . . . calls 'strong necessities'" is questionable.

Let us now see how Howell argues against CP′.[143] For starters, he assumes CP′ for reductio, and then claims that SN″$_1$ is coherently conceivable. From these two premises, it is a straightforward inference to the conclusion that SN″$_1$ is possible.

1.   CP′   [reductio assumption]

2.   SN″$_1$ is coherently conceivable.   [a priori]

3. $\lozenge$SN″$_1$ [1, 2]

From here, Howell aims to derive SN″. For he believes that SN″ implies $\neg$CP′, so that, by deriving SN″, he can complete his reductio against CP′. All in all, then, Howell argues that CP′, if true, is false. For CP′, if true, implies SN″, which implies $\neg$CP′.

---

[143] The following paraphrase is derived from Howell 2008, p. 352.

As we will see, Howell's derivation of SN″ is a two-step process: first he derives ◊SN″$_2$,

and then he derives SN″. Howell's defense of ◊SN″$_2$ relies heavily on his claim that SN″'s

"primary and secondary intensions are the same," i.e., that they "coincide" with respect to SN″.

What he means is that the 1- and 2-intensions are such that, with respect to SN″, they map all of

the same worlds to all of the same truth values.[144] For short,

4.   *The Coincide Hypothesis:* □(SN″$_1$≡SN″$_2$)

Howell defends *The Coincide Hypothesis* as follows:

> Why say that the two intensions must coincide? The reasoning . . . is that [SN″], if true, is
> necessary (since it is a statement about the whole of logical space) as well as a priori. (If
> it was a posteriori, one wonders how it could be empirically established. It is certainly
> treated as an a priori falsehood by the two-dimensionalists.) If a statement is necessary a
> priori, its truth across possible worlds considered as counterfactual and as actual should
> coincide, because its truth requires no contribution from any particular world. Primary
> and secondary intensions come apart in the water case, since which world is actual
> matters to the evaluation of necessary truths concerning water. That is why the necessities
> there are a posteriori. In the case of a priori truths, by hypothesis one doesn't need to find
> out which world is actual in order to evaluate them. Given the way primary and
> secondary intensions are defined, therefore, they should coincide when it comes to the
> necessary a priori. So, premise five follows from the modal and epistemological nature of
> SN and the definitions of primary and secondary intensions.[145]

To paraphrase:

i.   SN″ is either true or false. [bivalence]

ii.   SN″, if true, is necessarily true[146] and knowable a priori. ["it is a statement about

the whole of logical space"]

---

[144] Strictly, the 1- and 2-intensions coincide with respect to SN″ iff, for all x and y, if x is a centered world, and y is
the uncentered part of x, then SN″$_1$ is true in x iff SN″$_2$ is true in y.
[145] Howell 2008, p. 352-3.
[146] Here one might wonder what it should mean to say that SN″ is necessarily true. Ordinarily, to say that some
statement is necessarily true is to say that it is true in all possible worlds. Yet SN″ is a meta-modal claim about the
set-theoretic relationships between the space of conceivable scenarios and the space of logically possible worlds.
Thus one might argue that SN″'s necessity should require there to be a space of possible meta-modal claims about
the set-theoretic relationships between the space of conceivable scenarios and the space of logically possible worlds,
whereby SN″ is true of each of these meta-modal claims. In that case, one might argue that SN″'s being true in all
possible worlds doesn't guarantee SN″'s being necessarily true in the meta-modal sense just described; for being
true in all possible worlds is consistent with being false in at least one world that is impossible in our meta-modality,

iii.   For any statement, S, that is necessarily true and knowable a priori, S_1 and S_2 will both be necessarily true. [a priori necessary "truth requires no contribution from any particular world"]

iv.   So, if SN'' is true, then SN''_1 and SN''_2 are both necessarily true, in which case *The Coincide Hypothesis* is true.

v. SN'', if false, is necessarily false and its negation is knowable a priori. ["it is a statement about the whole of logical space"]

vi.   For any statement, S, that is necessarily false and whose negation is knowable a priori, S_1 and S_2 will both be necessarily false. [a priori necessary falsity requires no contribution from any particular world]

vii.   So, if SN'' is false, then SN''_1 and SN''_2 are both necessarily false, in which case *The Coincide Hypothesis* is true.

viii.   So, *The Coincide Hypothesis* is true. [i, iv, vii disjunction elimination]

From here, Howell argues that $\lozenge SN''_1$ and *The Coincide Hypothesis* together imply $\lozenge SN''_2$. For $\lozenge SN''_1$ says that SN''_1 is true in some possible worlds, while *The Coincide Hypothesis* implies that SN''_2 is also true in any such world; thus SN''_2 is possible as well.

5. $\lozenge SN''_2$

Howell then argues that $\lozenge SN''_2$ implies SN'' on the grounds that, since SN''_2 is either necessarily true or necessarily false, SN''_2's possible truth entails its necessary truth, and thus its actual truth.

6. SN''

---

but not impossible in some other meta-modality. Yet I suspect that there is only one possible meta-modality, as I think the option of multiple different meta-modalities is incoherent. At any rate, it would require treating some possibilities as being possible only meta-contingently, but this doesn't coincide with what I (and most philosophers) would like to mean by a *possibility*. At any rate, taking the option of multiple meta-modalities seriously would require being committed to restrictions on meta-modal accessibility, similar to restrictions on modal accessibility, except that in this case the inaccessible worlds would, not be "far away" possible worlds in the same meta-modal space, but rather would be possible worlds in alternative meta-modal spaces. This requires too much arbitrariness, I think. If a world is possible in some meta-modal space, why wouldn't it be possible in every meta-modal space?

Howell then argues that SN″ implies ¬CP′. He writes:

> But if [SN″] is true, [CP′] is false. Only if [SN″] is false can we be guaranteed that there is a world corresponding to the truth of our primary intensions. If [SN″] is true, therefore, there is no entailment from the conceivability of the primary intension to its possibility.[147]

The argument here is that, since CP′ implies ¬SN″, SN″ implies ¬CP′.

     7.   SN″⊃¬CP′

Finally, by modus ponens from lines 6 and 7, we can infer ¬CP′.

     8.   ¬CP′

Thus, from the reductio assumption that CP′ is true, one can derive that CP′ is false. This is supposed to show the absurdity of CP′.

Before I evaluate this argument, let me first point out an oddity. Namely, it is odd that Howell bothers to derive ◊SN″$_2$ at all. I would have expected him to argue as follows: ◊SN″$_1$ implies □SN″$_1$, which implies SN″$_1$, which implies ¬CP′. In private correspondence with Howell, however, I learned two interesting things. First, he thought it was unnecessary to derive SN″$_2$ but decided to do so anyway. Second, he mistakenly took Chalmers to hold that any unqualified statement S is ambiguous between S$_1$ and S$_2$, apparently thinking that S$_2$ is in some sense the more "metaphysical" claim. Thus he decided to derive SN″$_2$ just in case the argument would look more compelling with ¬CP′ being derived from SN″$_2$.

Unfortunately, Howell's strategy involves a misunderstanding of Chalmers' two-dimensional semantics. For Chalmers makes clear in his response to Diego Marconi's (2005) paper, "Two-Dimensional Semantics and the Articulation Problem," that he rejects the "ambiguity" view according to which an unqualified statement S is ambiguous between S$_1$ and S$_2$; instead, Chalmers holds that S involves both intensions. He writes:

---

[147] Howell 2008, p. 353.

> My view is . . . [that] ordinary expressions have a complex semantic value (no ambiguity) involving both 1-intensions and 2-intensions. Different aspects of this semantic value play a role in determining a truth-value in different contexts: e.g. 'It is a priori that S' is true when S has a necessary 1-intension; 'It might have been that S' is true when S has a possible 2-intension. (Note that this does not entail that S is ambiguous, or that it has a context-dependent semantic value.)[148]

Thus, according to Chalmers, SN'' is not ambiguous between SN''_1 and SN''_2; rather, assuming that there are only two intensions, SN'' is equivalent to the conjunction, $SN''_1 \& SN''_2$. It follows that, if we want to derive SN'', we can do so by first deriving $SN''_1 \& SN''_2$, which is equivalent to SN''. For instance, one might argue to line 6 as follows:

5.1. $\Diamond SN''_1 \& \Diamond SN''_2$ [3, 5 & intro]

But the truth values of SN''_1 and SN''_2 are necessary, since (as Howell puts it) they are "about the whole of logical space." Thus their possible truth implies their necessary truth.

5.2. $\Box SN''_1 \& \Box SN''_2$ [5.1]

But their necessary truth implies their actual truth; so, we can derive SN'' as follows:

5.3. $SN''_1 \& SN''_2$ [5.2]

5.4. $SN''_{1,2}$ [5.3 equivalent]

6. SN'' [5.4 equivalent]

This way of deriving SN'' is in line with Chalmers' understanding of his two-dimensional semantics as a "complex semantic value" thesis instead of as an "ambiguity" thesis.

Aside from the above-mentioned infelicity, Howell's argument is fairly well reasoned; nevertheless I will argue that it is problematic and unsound.

The problem is that line 2 is highly contentious. For line 2 says that SN''_1 is coherently conceivable, and this seems pretty counterintuitive to me. At any rate, on most definitions of "coherent," it will not be coherently conceivable for a coherent conception to be logically

---

[148] http://consc.net/responses.html#marconi

79

impossible, which is what SN''_1 requires. Rather, a coherent conception's being logically impossible doesn't seem to make sense, and thus appears to be incoherent and therefore not coherently conceivable. Yet Howell doesn't address this problem. Fortunately, however, he does give some reasons that can be taken in support of the coherent conceivability of SN''_1. Particularly, he cites two potential sources of strong necessities in defense of the coherent conceivability of SN''_1.

The first potential source of strong necessities comes from Colin McGinn. McGinn argues that the mind-body problem might be unsolvable for the reason that discovering its solution would require conceiving of a property that is inconceivable.[149] The idea, then, is that it may be coherently conceivable that McGinn's proposal is true in such a way that implies SN''_1. If this is coherently conceivable, then SN''_1 is coherently conceivable.

Unfortunately, this argument is not as strong as it may seem at first sight. For one thing, if we distinguish between coherent *positive* conceivability and coherent *negative* conceivability, we'll see that McGinn's proposal cannot be coherently positively conceivable, as that would require positively conceiving the inconceivable properties, which is impossible. Thus McGinn's proposal is at best only coherently negatively conceivable, since this merely requires a conception that fails to involve the noticing of any contradiction implied by the conceived state of affairs. On this horn, the conception would merely imply that SN''_1 is coherently negatively conceivable, in which case deriving $\Diamond SN''_1$ would require the following principle:

CP′- Coherent primary negative conceivability entails primary possibility.

Thus, by citing the example from McGinn, Howell thereby gives an example that can at best be used in a reductio against CP′-; it cannot work in a reductio against the positive analog, CP′+:

CP′+ Coherent primary positive conceivability entails primary possibility.

---

[149] McGinn's classic statement is his (1989) paper, "Can We Solve the Mind-Body Problem."

More generally, no CP principle that is run in terms of positive conceivability is threatened by McGinn's proposal. This is in addition to the fact that, although McGinn's proposal is difficult or impossible to refute, it is highly speculative and it is difficult or impossible to make it seem plausible.

The other potential source of strong necessities that Howell cites is Brian Loar's (1997) claim that, "due to the distinct functions of phenomenal concepts and scientific concepts they can have the same extensions even if we cannot realize that fact *a priori*."[150] The idea is that, for all we know, the physical and the phenomenal might be two aspects of the same underlying reality. Specifically, Loar considers the option that phenomenal and scientific concepts are co-extensive in all logically possible worlds, in which case the options of qualia inversion and zombies are logically impossible. In such a case, there will be a counterexample to CP′ if the option of qualia inversion or zombies is coherently conceivable. For, where S asserts the presence of qualia inversion or zombies, S_1 will be coherently conceivable yet logically impossible, thereby serving as a counterexample to CP′, in which case CP′ is false.

This is a serious problem for Chalmers. For there doesn't seem to be any way to prove the logical possibility of qualia inversion or zombies. With that said, it's rare to find a philosopher who denies the logical possibility of qualia inversion or zombies. For even materialists tend to accept the logical possibility of qualia inversion or zombies; as a rule, they are suspicious of the *metaphysical* possibility, but not of the logical possibility. Thus, although Loar's proposal is serious, it is also highly contentious.

In sum, McGinn's proposal is difficult to refute but also difficult to make plausible, and it only threatens CP principles that are defined in terms of negative conceivability. Meanwhile, Loar's proposal that qualia inversion and zombies are logical impossibilities is a serious problem

---

[150] Howell 2008, p. 353.

for CP′, although it is also highly contentious, as many philosophers (including most materialists) reject the proposal. Thus Howell's justification for the coherent conceivability of SN''_1 is a bit cursory. Not only doesn't he address my critique of the coherent conceivability of SN''_1, but each of the examples that he cites as potential sources of strong necessities is problematic in its own right.

Near the end of his paper, Howell writes: "Chalmers will presumably deny . . . that $[SN''_1]$ is ideally conceivable."[151] That is, Chalmers might ignore SN''_1, which is written in terms of coherent conceivability, and instead focus on denying the following thesis, which is written in terms of ideal conceivability:

SN‴   For some S, S_1 is ideally conceivable but not logically possible.

SN‴ is equivalent to SN, which Chalmers denies. Yet Howell says that one can take this line only at the cost of compromising two-dimensionalism in two other ways.

First, insisting on incorporating the qualification of *ideal* conceivability requires a notion of ideal conceivability (as opposed to mere coherent conceivability), which Howell "suspects will receive a rather circular definition: conceivability is ideal iff it tracks possibility."[152] He says that this would replace "the old gap between conceivability and possibility with a new gap between prima facie conceivability and ideal conceivability,"[153] and that this new gap "seems every bit as substantive as the original inference from conceivability to possibility."[154]

Second, insisting on incorporating the qualification of *ideal* conceivability is further problematic in so far as it involves asserting that, in the case of SN'''_1, "we cannot grasp it well

---

[151] Howell 2008, p. 354.
[152] Howell 2008, p. 354.
[153] Howell 2008, p. 354.
[154] Howell 2008, p. 354.

enough to form a robust conception of its truth or falsity."[155] He compares it to how a two-dimensionalist *might* handle Goldbach's conjecture or the possibility of a necessarily existing God. He says that the two-dimensionalist might hold that these theses cannot be understood well enough to be justifiably accepted or rejected. Thus, *if* the same attitude should be held regarding SN'''_1, then SN'''_1 too shouldn't be accepted or rejected, in which case, "we should not employ a system [e.g., a two-dimensional system] that presupposes $[SN'''_1]$'s falsity."[156] Finally, Howell says that these two reasons demonstrate that, "even if the metaphysical two-dimensionalist avoids a straightforward *reductio*, he does so by admitting that two-dimensionalism is too shaky a tool to be used."[157]

On the first point, Howell is mistaken in two respects. First, Chalmers doesn't define conceivability as ideal iff it tracks possibility. Indeed, in his (2002a) paper, he says, "it is important that 'better reasoning' about conceivability not be defined even in part as reasoning that better tracks possibility. Such a criterion would trivialize the link between ideal conceivability and possibility."[158] Rather, Chalmers defines conceivability as ideal iff it lacks cognitive defect. If it happens to track possibility, this is supposed to be a substantive fact, not due to a circularity of definition. Furthermore, it is infelicitous for Howell to say that this new gap between prima facie conceivability and ideal conceivability would be "every bit as substantive" as the old gap between conceivability and possibility. The old gap was more substantive, at least when construed in terms of MD. When MD is assumed, conceivability is a poor guide to metaphysical possibility, as, even if conceivability is a reliable guide to logical possibility, some logical possibilities are ruled out as metaphysically impossible in ways that

---

[155] Howell 2008, p. 355.
[156] Howell 2008, p. 355.
[157] Howell 2008, p. 355.
[158] Chalmers 2002a, § "Prima Facie vs. Ideal Conceivability."

seem rationally arbitrary. Yet when MD is not assumed, the most constrained space of worlds is

the space of logically possible worlds, which is only constrained in logical ways, and thus it is

less likely to be constrained in ways that should seem rationally arbitrary.

On the second point, Howell's hypotheticals are all true. The only problem is that their

antecedents are not all true. In effect, he is suggesting that two-dimensionalism has commitments

which it doesn't in fact have. For instance, Chalmers does *not* hold that SN'''_1 cannot be grasped

well enough for us to form a robust conception of its truth or falsity. Indeed, he argues that SN is

false, which is equivalent to SN'''. This is in addition to the fact that a two-dimensionalist *doesn't*

need to hold that Goldbach's conjecture or the necessarily existing God hypothesis is

insufficiently grasped. Indeed, I discussed earlier in this chapter how the Goldbach conjecture

and the necessarily existing God hypothesis might be handled without resorting to saying that

they are insufficiently grasped.

In sum, Howell gives a reductio argument that CP′ implies a contradiction. line 2

contentiously assumes the coherent conceivability of SN'''_1, which I argued against. I also

discussed Howell's two considerations in favor of the coherent conceivability of strong

necessities, which he takes as justification for the coherent conceivability of SN'''_1. The first

proposal (McGinn's) is highly contentious and can only be a problem for *negative*

conceivability-to-possibility principles. Meanwhile, the other proposal (Loar's) is also highly

contentious, so that altogether Howell's justification for the coherent conceivability of SN'''_1 is a

bit cursory. Finally, Howell anticipates that a two-dimensionalist might reply to his reductio

argument by holding that SN'''_1 is not ideally conceivable. Yet he gives two reasons to think that,

if the two-dimensionalist responds in this way, two-dimensionalism will be compromised. As we

have seen, however, Howell's two reasons aren't as powerful as he thinks. The first contains two

mistakes, and the second employs false assumptions about what two-dimensionalists are committed to. Thus Howell's critique of Chalmers' two-dimensional argument against materialism is problematic in several ways.

Is CP- worth employing?

I have discussed several objections to Chalmers' CP principles, some of which are objections to CP-. I will now argue, however, that ideal negative conceivability is incoherent and thus CP- is not worth employing, so that Chalmers' two-dimensional argument against materialism should instead be run in terms of CP+.

Recall that negative conceivability is supposed to be a kind of failure: namely, a failure to rule out the truth of a given statement by a priori means. As Chalmers puts it: "The central sort of negative conceivability holds that S is negatively conceivable when S is not ruled out a priori, or when there is no (apparent) contradiction in S."[159] It follows that, for any statement, S, I negatively conceive of S when I am in an unthinking coma. For, since I fail to engage in any a priori thought whatsoever, I fail to rule out S a priori and thus meet the sufficient condition of negatively conceiving of S. More generally, then, in so far as negatively conceiving is simply the failure to rule out a given statement by a priori means, then, for every statement, S, anything that doesn't think at all negatively conceives of S.

As odd as this is, it is not a sufficient reason to reject negative conceivability as a useful concept in philosophy. For, despite the oddity that unthinking beings would count as negatively conceiving of every statement, it may be possible to qualify negative conceivability in such a way that would render the qualified form useful in philosophy. Particularly, perhaps *ideal negative conceivability* can be a useful concept in philosophy. Indeed, Chalmers treats it as such. But what would it take to make a non-ideal negative conception so much better that it would be

---

[159] Chalmers 2002a, § 2, second paragraph.

an ideal negative conception? Presumably it would require that, in addition to a failure, there is also a positive aspect. Unfortunately, it seems that the only way for a conception to incorporate something positive would be for it to be a *positive conception* in Chalmers' sense of involving some sort of image or proof. Yet this would mean that, for a negative conception to be ideal, it would have to be a positive conception after all.

Yet that is not all. For being a positive conception that meets the negative condition of failing to rule out the statement's truth a priori is still not sufficient for being *ideal*. To be ideal, the positive component has to be of high enough quality. Otherwise (say) my cat would ideally negatively conceive of a box that has exactly five and six sides simply by experiencing redly in conjunction with a failure to rule out that statement's truth. But presumably the only way for the positive component's quality to be sufficiently high as to render the conception ideal is if the conception were an *ideal positive conception* in the first place.

Thus "ideal negative conceivability" is a misnomer. For the phenomena in question is not a kind of negative conceivability after all, but is ideal positive conceivability in disguise. Consequently, Chalmers' two-dimensional argument against materialism should be written in terms of ideal positive conceivability and should appeal to CP+. Thus any objection to CP- that is not also an objection to CP+ should be ignored as irrelevant to the way that Chalmers' argument should be run. Thus we should ignore some of the objections I discussed earlier—e.g., objections from strong mathematical necessities or objections from strong necessities clouded by epistemic vagueness—as irrelevant to the way that Chalmers' argument should be run, since they don't threaten CP+.

CHAPTER 5

IS CHALMERS' CONCEIVABILITY PREMISE TRUE?

In this chapter, I discuss criticisms of Chalmers' *Conceivability Premise*—the premise that $P_1T_1\&\neg Q_1$ is ideally positively conceivable. First, I discuss Stephan Leuenberger's (2008) blocker argument which holds that the *Conceivability Premise* might be false since its "that's all" clause might render it not positively conceivable. Second, I discuss Keith Frankish's (2007) "anti-zombie" argument which holds that the justification for the *Conceivability Premise* and CP principles is undermined due to the fact that an analogous, equally good argument can be run in defense of the conclusion that materialism is true.

Leuenberger's blocker argument

Stephan Leuenberger (2008) argues that Chalmers' *Conceivability Premise* might be false on the grounds that its "that's all" clause might not be positively conceivable. The reason he thinks it might not be positively conceivable is that it might require positively conceiving the absences of certain beings or properties that are so unfamiliar to us that we cannot positively conceive of their absence.

Particularly, he assumes that there might be possible *blockers* whose absences we cannot positively conceive. The "blocker" terminology comes from John Hawthorne's (2002) paper, "Blocking Definitions of Materialism." For Hawthorne and Leuenberger, a blocker is any being or property that prevents a mental property from being instantiated that would otherwise have been instantiated given the physical properties. That is, for any physical property, p, and mental property, m, b is a blocker for m relative to p iff $p\&\neg b$ entails m, but $p\&b$ entails $\neg m$.

For instance, my current brain state might be sufficient for my current visual state in our world, but not in worlds that have extra ingredients that prevent the supervenience. Such ingredients would thereby serve as blockers. In conjunction with my brain state, they would yield blindness, say, or some difference in spatial representation or color.

The most troublesome kind of blocker, according to Leuenberger, would be both non-physical and non-mental, as this would make them exceedingly unfamiliar to us and therefore incomprehensible to us. According to Leuenberger, this would make us incapable of positively conceiving of their presence or even their absence. For he "denies that we can positively conceive the absence of facts that are both positive and unfamiliar."[160] The problem this poses for Chalmers' *Conceivability Premise* is that, if non-physical, non-mental blockers are possible, then the "that's all" clause of Chalmers' *Conceivability Premise*,

$P_1T_1\&\neg Q_1$ is ideally positively conceivable.

will imply that there are no non-physical, non-mental blockers. For the *Conceivability Premise* will thus claim that the absence of non-mental, non-physical blockers can be positively conceived, which Leuenberger denies. Thus Leuenberger concludes that, for all we know, the *Conceivability Premise* might be false, since it is false if non-physical, non-mental blockers are possible, and we have no reason to reject the possibility of non-physical, non-mental blockers.

The contentious premise is Leuenberger's claim that absences of unfamiliar facts are not positively conceivable. For it is unclear why degree of familiarity should have anything to do with whether a given *absence* is positively conceivable. It is clear why degree of familiarity should make a difference to whether a thing's *presence* can be positively conceived, as this requires imagining its existence; but why should we have to be familiar with a being or property in order to imagine its absence? Indeed, Chalmers replies to Leuenberger by saying: "Absences

---

[160] See Leuenberger 2008, p. 21.

are positively conceivable, and absences of unfamiliar entities (aliens, ghosts, blockers) are not

different in principle. Leuenberger does not give a reason to believe that if we were familiar with

blockers, the situation would be any different."[161]

Chalmers is right that Leuenberger "does not give a reason to believe that if we were

familiar with blockers, the situation would be any different." However, Leuenberger does hint at

some reasons in the form of a clarification. For first he gives a formal statement of his view that

positive facts are positively conceivable, but negative facts are positively conceivable only when

they "concern properties of our acquaintance . . . [i.e., only when they involve] properties and

relations about which we have *de re* thought. . . . [i.e., only when they are] *familiar*."[162] His

"informal" clarification of this sheds light on what his reasons are for accepting his key premise:

> Informally, the idea is that in imagination, whatever is not specified by you does not get
> specified by default. In particular, the negation of a fact or the absence of a property does
> not get specified by default. The limiting case is illuminating: if you do not imagine
> anything, you do not count as imagining that there is nothing. Unless you make the
> mental image very specific, what the image represents is indeterminate between many
> different worlds. With respect to facts that you are not acquainted with, you are not in a
> position to make the image specific.[163]

Here Leuenberger is commenting on the nature of imagination. He holds that, "in imagination,

whatever is not specified by you does not get specified by default." On the face of it, this has an

air of plausibility. For instance, it's not a property of imagination that, in imagination, if one does

not specify how many birds there are, the default representation is *seven* birds—so that, despite

imaging no birds, the imagination nevertheless *represents* seven birds. Nor is the default to

represent exactly *zero* of whatever isn't imagined. For that would require positive conceptions to

be of total worlds. But positive conceptions do not have to be of total worlds. As Leuenberger

puts it, "It is illegitimate to move from 'I am positively conceiving a situation in which A' to 'I

---

[161] Chalmers 2010, p. 164.
[162] Leuenberger 2008, p. 21.
[163] Leuenberger 2008, p. 21.

am positively conceiving the world completely described by A.' It is not part of the content of the positive conception that the imagined situation is not merely a partial, but a total world."[164] So, on the face of it, Leuenberger appears to be right that, in imagination, nothing gets specified by default.

Yet, if nothing in imagination gets specified by default, it's difficult to see how negative facts might get represented in imagination. For how might one imagine the *lack* of a given property? The problem is that imagination seems positive (about what is present), whereas absences are by definition negative (about what is not present). Thus imagining absences seems paradoxical at best. Indeed, Leuenberger's initial hunch was to deny the possibility of positively conceiving absences. For he says he has "an 'atomist' conception of positive conceivability, according to which we positively conceive complex facts only in virtue of conceiving atomic facts,"[165] and in his preliminary formulation of his inconceivability principle he holds that positively conceiving absences is impossible.[166]

Nevertheless Leuenberger accepts that absences are positively conceivable on the grounds that it seems clear that we can positively conceive, say, a table that doesn't have any apples on it.[167] He figures, however, that the only reason we can positively conceive the absence of apples is that we are familiar with apples. Although he doesn't explain why he thinks our ability to imagine the absence of apples is due partly to their familiarity, his idea is presumably that, in imagination, we can depict absences of certain things, but only by forming images of them in the right way—for instance, by forming images of them, and then crossing them out, fading them out, or making them vanish so as to have some positive way of representing *their*

---

[164] Leuenberger 2008, p. 18.
[165] Leuenberger 2008, p. 20.
[166] See Leuenberger 2008, p. 20-1.
[167] See Leuenberger 2008, p. 21.

absence. By contrast, it is unclear how a person might positively represent the absence of unfamiliar properties in the same sort of way. For, since we wouldn't know how to depict unfamiliar properties, we wouldn't know how to cross them out, fade them out, make them vanish, or do anything to those depictions that would require having those depictions in the first place.

Another way that we might try to positively conceive absences is by doing what Leuenberger calls "annotatedly conceiving." In his words, this would be to "accompany a mental image with a commentary, e.g. 'no other positive properties are instantiated.'" This option has the advantage that it wouldn't require that, in imagination, indeterminacies are removed *by default*. Yet Leuenberger points out that annotatedly conceiving is nevertheless problematic, as it is not as good as positively conceiving. As Leuenberger puts it, "the Conceivability Argument with 'annotatedly conceive' in the place of '[positively] conceive' is not successful. Unless there are restrictions on what the commentary can be, annotated conceivability does not imply possibility."[168] Leuenberger backs this up with some examples. In one example, the added commentary is itself impossible. Namely, "I can positively conceive that Thomas Hobbes is drawing figures and writing symbols on a piece of paper, and add the commentary 'Hobbes squares the circle.'"[169] In another example, the added commentary is not impossible, but it contradicts what was depicted in imagination. Namely, "I can positively conceive that a particle travels at 500 million meters per second, and add the commentary 'The particle obeys the laws of Special Relativity.'"[170] So, Leuenberger is right that, if annotated conceivability is going to be a reliable guide to possibility, there will have to be "restrictions on what the commentary can be." The only possible restriction that he considers is the restriction that the commentary must be

---

[168] Leuenberger 2008, p. 22.
[169] Leuenberger 2008, p. 22.
[170] Leuenberger 2008, p. 22.

consistent with what is positively conceived.[171] However, he points out that this option is a non-starter, as it would trivialize the relevant kind of CP principle. Namely, one would have to defend a hybrid conceivability-to-possibility principle, such as follows:

CP+$_{an}$ For any S, if a portion of S_1 is ideally positively conceivable, and the rest is an annotated conception that is consistent with the former, then $\Diamond S_1$.

The problem with a CP+$_{an}$ is that it's *trivially* true. For to say the annotated conception . . . is *consistent* with the ideally conceivable component is to say that there's a possible world where both components obtain. (Consistency implies possibility.) Thus the possibility that one aims to derive is already presupposed by the hybrid positive/annotated conceivability in question. Thus, in any given case, to say that one's conception is conceivable in this way while also saying that CP+$_{an}$ is true will be question-begging. Nevertheless, besides saying that the annotated commentary must be consistent with the ideally positively conceived component, Leuenberger writes: "I cannot see what other restriction would ensure that it implies possibility."[172]

In sum, Leuenberger's argument appears to have two components. First, he appears to implicitly think that positively conceiving absences involves depicting absent properties in ways that resemble them—say, depicting them and then crossing them out, or making them fade or vanish. Yet he holds that this is a process that we cannot do with unfamiliar properties, since we don't know how to depict unfamiliar properties via images that resemble them. Second, he thinks that the only alternative to the way that we positively conceive absences of familiar properties is by *annotatedly* conceiving them. Yet he suspects that there is no way of restricting annotated conceivability that would render ideally positive conceivability (with an annotated commentary

<hr>

[171] Leuenberger 2008, p. 22.
[172] Leuenberger 2008, p. 22.

component) a non-trivially reliable guide to possibility. Nevertheless, there are weaknesses with both components of Leuenberger's argument, which I will now address.

First, although it is true that we cannot depict unfamiliar properties in ways that resemble them, I argue that we should not have to do this in the case of *universal* "that's all" clauses (= "that's all" clauses that claim that *there are no other fundamental beings or properties of any kind whatsoever*.) Rather, we can quantify over every other (familiar or unfamiliar) possible fundamental being or property with a symbol that doesn't resemble them, and then cross it off, or make it fade or vanish. In this way, we will positively conceive of their absence. Moreover, there will be no danger of indeterminacy; for, since universal "that's all" clauses claim that there are no other fundamental beings or properties of any kind, it cannot be left indeterminate which kinds of unfamiliar fundamental beings or properties are thereby said to be absent. They will *all* be taken as absent.

Yet the "that's all" clause in Chalmers' *Conceivability Premise* is a universal "that's all" clause; for it says that *there are no other basic positive truths*, and thus no other fundamental beings or properties of any kind, familiar or unfamiliar. Thus there's no threat of indeterminacy in the positive conceivability of Chalmers' *Conceivability Premise* that can be rightly blamed on its "that's all" clause. For the mental act of positively conceiving its "that's all" component can be performed by considering a symbol that represents (without resembling) every other possible fundamental being or property that might contribute extra structure, and then crossing off this symbol, or making it fade or vanish so as to represent the absence of any extra structure, and thus the absence of any extra fundamental beings or properties.

Second, Leuenberger overlooks a plausible way that positive conceivability (with an annotated conceiving component) might be a reliable guide to possibility without trivializing the

entailment. Namely, we can simply restrict the annotatedly conceiving component to universal "that's all" clauses. This way, instead of citing CP+, Chalmers can cite,

> CP+*   Ideal primary positive conceivability (with or without an annotatedly conceived universal "that's all" clause) entails primary possibility.

This way there is no chance that a relevant annotated component (i.e., a universal "that's all" clause) will yield indeterminacy in the conception, since there is nothing indeterminate about ruling out *every other possible fundamental being or property*. There is also no chance that a relevant annotated component will be self-contradictory, as absences cannot be self-contradictory. Thus the only potential problem is that a relevant annotated component might be inconsistent with certain positively conceived scenarios. For instance, one might positively conceive of a quark and that's all. That is, a universe that consists entirely of a single quark. Yet this might be impossible: e.g., it might be necessary that (as in our world) quarks only come in trios. At any rate, if it is impossible for there to be a single quark and that's all, then the universal "that's all" clause will be inconsistent with the claim that there is a single quark. Thus, on the face of it, there is a danger that CP+* won't be as plausible as CP+, since CP+* implies that tacking on a universal "that's all" clause will not decrease the reliability of tracking possibility. Fortunately, however, this isn't a problem for Chalmers' *Conceivability Premise*. After all, we already know from the possibility of our world that our world's structure can exist without there being any additional structure, and thus without there being any additional possible fundamental beings or properties. For our world is itself a case of a world that has our world's structure without having any additional structure.

Altogether, there are two weaknesses with Leuenberger's argument, corresponding to its two components. First, it seems that absences of unfamiliar properties can be positively

conceived by entertaining a symbol that represents them (without resembling them), and then crossing this symbol off, or making it fade or vanish so as to positively conceive the absence of any such property. By doing so with a universal "that's all" clause, the absence of additional fundamental beings or properties will be positively conceived without any of the indeterminacy that Leuenberger warns about. Second, even if we grant that absences of unfamiliar properties cannot be positively conceived, Chalmers' *Conceivability Premise* can nevertheless be justified by CP+* instead of CP+, and running the argument with CP+* instead of CP+ does not appear to introduce any new problems. For instance, unlike CP+$_{an,}$ CP+* is not trivially true; for it doesn't explicitly require that the annotated commentary be *consistent* with the ideally positively conceived component in order to entail possibility. Furthermore, there's no threat in Chalmers' argument that the annotated component might be inconsistent with the ideally positively conceivable component. For it is clear there's a possible world that has our world's structure without having any additional structure, since our world is such a world. So, Leuenberger's argument against Chalmers' *Conceivability Premise* is misguided and has several implausible aspects.

Now let us discuss another critique of Chalmers' *Conceivability Premise*. Unlike Leuenberger's argument that Chalmers' *Conceivability Premise* is impossible, Keith Frankish merely argues that it is insufficiently justified.

Frankish's "anti-zombie" argument

Keith Frankish's (2007) "anti-zombie" argument is a parity of reasons argument. Parity of reasons arguments are popular in the history of philosophy. They are a kind of argument by analogy where an argument is attacked by giving an analogous argument that is somehow implausible, or unsound, or whose conclusion is inconsistent with the conclusion of the target

argument. The most well-known parity of reasons argument is Guanilo's parody of Anselm's ontological argument which aims to "prove" the existence of the greatest conceivable island.

Frankish's parity of reasons strategy is to hold that, whereas Chalmers gives a *zombie* argument against materialism, one can just as well give an "anti-zombie" argument for materialism that is about equally plausible. To be clear, Frankish is not particularly concerned to defend materialism. His strategy is rather to argue that, since both materialism and anti-materialism can be defended in the same sort of way, with about equal plausibility, the argument form that they both employ should not be taken seriously. Indeed, he argues that there are two main weaknesses inherent in the form: the conceivability premise and CP.[173]

Before I go over Frankish's argument, let me point out a couple of oddities. First, regarding the two-dimensional nature of Chalmers' argument, Frankish says that he is going to ignore it. As he says: "[Chalmers'] argument is often set out using the framework of two-dimensional semantics, but I shall present it in a simpler version here; nothing will turn on the complications omitted."[174] This is alarming, as incorporating the two-dimensional semantics is supposed to make a big difference with respect to the seriousness of its conceivability problems: for when two-dimensional semantics is employed, both the *Conceivability Premise* and the inference to the *Possibility Premise* are supposed to be more plausible. Fortunately, after saying that he is going to ignore the two-dimensional component, Frankish nevertheless goes on to qualify the conceivability as "ideal and primary":[175] the qualification of *primary* qualifies it as being evaluated by 1-intensions, which is part of the two-dimensional framework. Yet in this passage he only applies the qualification of primary to the conceivability, not to the possibility.

---

[173] Brown (2010, 2012) defends pretty much the same argument, except that he assumes CP and argues that the *Conceivability Premise* is false.
[174] Frankish 2007, p. 651.
[175] Frankish 2007, p. 652.

He then defines "CP" as "the thesis that if a situation is ideally conceivable, then it is metaphysically possible."[176] This is a misleading way to put it, as often when people say "metaphysically possible" they intend a kind of modal dualism where the space of metaphysically possible worlds is smaller than the space of logically possible worlds. Likewise, by "metaphysically possible" Chalmers himself generally means *logically possible when evaluated via 2-intensions*. However, if we are to read Frankish charitably, we will have to read him as saying that, according to Chalmers, CP is the thesis that ideal primary conceivability entails *primary possibility*, so that, if it is ideally conceivable when evaluated by 1-intensions, then it is possible when evaluated by 1-intensions.

A second oddity is that Frankish takes himself to be arguing against an argument from Chalmers whose conclusion is that materialism is false. This doesn't do justice to the fact that Chalmers thinks it is important to defend the inclusive disjunction that *materialism is false or Russellian monism* is true in order to have less serious conceivability problems. Thus Frankish's straw man formulation runs the risk of exaggerating the very weaknesses that he wishes to exploit: for his agenda is to argue that Chalmers' argument has serious conceivability problems.

To see how Frankish's anti-zombie argument goes, we will have to go over the notion of an "anti-zombie." Roughly, whereas a *zombie* would be just like a conscious person physically except without being conscious, an *anti-zombie* would be just like a conscious person physically and consciously except that it would be entirely physical—that is, a "bare physical duplicate" of us,[177] having "no further properties of a non-physical kind."[178]

Strictly, Frankish defines anti-zombies, not only in terms of their intrinsic properties, but partly also in terms of what the rest of the world is like:

---

[176] Frankish 2007, p. 652.
[177] Frankish 2007, p. 653.
[178] Frankish 2007, p. 653.

> We can define anti-zombies as beings which are bare physical duplicates of us, inhabiting a universe which is a bare physical duplicate of ours, but none the less having exactly the same conscious experiences as we do. That is, in the anti-zombie world consciousness is a physical phenomenon, supervening metaphysically on the world's microphysical features – in virtue of token identities, say.[179]

By this definition, anti-zombies are "beings which are bare physical duplicates of us . . . but none the less having exactly the same conscious experiences as we do." In addition, they must inhabit "a universe which is a bare physical duplicate of ours." With this latter qualification, Frankish defines anti-zombies as inhabitants of worlds that are entirely physical. Presumably the reason he defines anti-zombies in terms of *worlds* is because his argument aims to challenge the view that materialism is false of our world, not merely false of our world's consciousness.

Frankish outlines his anti-zombie argument as follows:

1. Anti-zombies are conceivable
2. If anti-zombies are conceivable, then anti-zombies are possible
3. If anti-zombies are possible, then consciousness is physical
4. So consciousness is physical.[180]

Line 1 is the anti-zombie *conceivability premise*. It says that it is conceivable for there to be anti-zombies. Since Frankish holds that the conceivability is supposed to be qualified as *ideal* and *primary*, we can read line 1 as saying that anti-zombies are ideally 1-conceivable. Or, to cash out the option of anti-zombies as a description, line 1 can be re-written as follows:

1′. It is ideally 1-conceivable that, corresponding to our world which has conscious experiences, there is a possible bare physical duplicate world that has these same conscious experiences.

By qualifying this premise with a 1-intension, the phrase "bare physical duplicate" may seem to take on a strange meaning. For a bare physical duplicate is supposed to duplicate the physical

---

[179] Frankish 2007, p. 653.
[180] Frankish 2007, p. 654; Brown (2010) gives a similar argument.

properties while "having no further properties of a non-physical kind." Thus, when qualified with a 1-intension, this will amount to duplicating the structural/dynamical properties while having no further properties besides its structural/dynamical properties. This would require such a being's conscious experiences to be identical to structural/dynamical properties. Yet this would be a kind of eliminativism that denies their phenomenal character. However, Frankish's purpose is not to argue that conscious experiences don't have any phenomenal character, but merely to argue that their phenomenal character is physical. It follows, then, that Frankish has not formulated his conceivability premise in a way that does justice to his intended meaning. For the "that's all" function denoted by the qualification of being a *bare* physical duplicate is not supposed to indicate that an anti-zombie cannot have any properties besides structural/dynamical properties (i.e., physical$_1$ properties); it is merely supposed to indicate that an anti-zombie cannot have any properties besides its physical structural/dynamical properties or intrinsic natures. Indeed, Frankish makes clear later in the argument that anti-zombies can have intrinsic natures; for he writes: "This structure is metaphysically sufficient for consciousness, whatever the intrinsic nature of the microphysical properties involved."[181]

To accommodate Frankish's intended meaning, line 1′ will have to be re-written with both 1- and 2-intensions.

1″. It is ideally 1,2-conceivable that, corresponding to our world which has conscious experiences, there is a possible bare physical duplicate world that has these same experiences.

Line 2 says that, if anti-zombies are 1,2-conceivable, then they are 1,2-possible. Thus line 2 can be re-written as follows:

---

[181] See Frankish 2007, p. 657.

2′.   If it is ideally 1,2-conceivable that, corresponding to our world which has

conscious experiences, there is a possible bare physical duplicate of our world

that has the same conscious experiences as our world, then it is 1,2-possible that

there is a bare physical duplicate of our world that has the same conscious

experiences as our world.

The justification here is supposed to be the first line together with one of Chalmers' CP

principles, which will hold that ideal 1-conceivability entails 1-possibility. Yet it is clear that this

won't work, as lines 1″ and 2′ incorporate more than just 1-intensions, and thus Chalmers' CP

principles are the wrong tools for the job. I will come back to this shortly.

Line 3 says, "If anti-zombies are possible, then consciousness is physical." Its antecedent

is supposed to be the consequent of the previous line, so we should re-write line 3 as follows:

3′.   If it is 1,2-possible that there is a bare physical duplicate of our world that has the

same conscious experiences as our world, then consciousness is physical.

Frankish says that this line "requires explanation, but . . . should be uncontroversial."[182] His

explanation is to say that, since the microphysical facts are metaphysically sufficient for

consciousness in the anti-zombie world, they should be metaphysically sufficient for that same

consciousness in any world that has the same microphysical facts.

> In the anti-zombie world consciousness is physical, so the microphysical features of that
> world (the laws and distribution of properties) are metaphysically sufficient for
> consciousness, and any world with the same microphysical features will have the same
> distribution of phenomenal properties. But by definition our world has the same
> microphysical features as the anti-zombie one. Hence the microphysical features of our
> world are metaphysically sufficient for the existence of consciousness.[183]

Frankish appears to use "metaphysically sufficient" to indicate a *metaphysical entailment*, but he

does not articulate a reason to accept the entailment. Thus line 3′ needs a more rigorous defense.

---

[182] Frankish 2007, p. 654.
[183] Frankish 2007, p. 654.

Frankish hints at a more rigorous defense of line 3′ when defining anti-zombies. He says, "in the anti-zombie world consciousness is a physical phenomenon, supervening metaphysically on the world's microphysical features – in virtue of token identities, say."[184] The key phrase here is "token identities," as token identities can support the relation of metaphysical supervenience.

One way for there to be token identities is if there are type identities. Type identities are identities between property types. The idea is that, for any properties m and p, if m is type identical to p, then necessarily any instance of m is an instance of p, and vice versa. Thus, if mental property m is identical to physical property p, then any token of m will be identical to a token of p, and vice versa. Thus maybe some kind of toothache is type identical to some kind of C-fiber excitation. In that case, necessarily, every instance of that kind of toothache is identical to an instance of that kind of C-fiber excitation, and vice versa. So, if all of our world's conscious experiences are instances of types that are identical to physical types, then any possible world that duplicates our world's physical events will instantiate those physical types, and thus instantiate all of our world's conscious experiences. In this way, our world's conscious experiences can metaphysically supervene on our worlds physical events due to token identities.

However, usually when philosophers speak of token identities, they have in mind a kind of functionalism. Token identity theory in this sense identifies mental events as being the performers of certain roles. Thus to say that a token of m is identical to a token of p in this sense is to say that one plays the role specified by the other. Thus if being a p is defined as being one who makes a sale for Primerica, then anything that makes a sale for Primerica thereby counts as being a p. Thus, if I perform a sale for Primerica, then I am token identical with the p who made that sale. This kind of token identity theory would be true of all conscious experiences iff all conscious experiences perform the right physical roles. Moreover, if this kind of token identity

---

[184] Frankish 2007, p. 653.

theory were true, it might ground a relation of metaphysical supervenience between the physical

truths and the conscious experiences. For, in so far as our world's physical truths rigidly

designate our world's conscious experiences, they entail those conscious experiences. With that

said, in so far as our world's physical truths do not rigidly designate our world's conscious

experiences, but instead designate physical roles that designate different role-players in different

possible worlds, then they do not entail our world's conscious experiences. Thus, functionalist

token identity theory is *consistent* with, but doesn't guarantee, the option that our world's

conscious experiences metaphysically supervene on our world's physical truths.

Thus we have seen how token identities might ground a relation of metaphysical

supervenience between our world's physical truths and our world's conscious experiences. Thus

Frankish is right that, given the right token identities between our world's physical truths and our

world's conscious experiences, line 3′ will be true.

Line 4 concludes, "Consciousness is physical." This follows by two applications of

modus ponens from lines 1″, 2′, and 3′. As Frankish describes it, it expresses a relation between

our world's physical structure and our world's conscious experiences.

> This structure is metaphysically sufficient for consciousness, whatever the intrinsic
> nature of the microphysical properties involved.[185]

Frankish hereby concludes that our world's structure is metaphysically sufficient for our world's

conscious experiences, regardless of what kind of intrinsic natures are involved.

Again, Frankish takes the anti-zombie argument to function as a parity of reasons

argument against Chalmers' argument. The idea is that, since the anti-zombie argument and

Chalmers' zombie argument have the same valid form but their conclusions are mutually

inconsistent, at least one of the arguments must employ a false premise. But their only

---

[185] See Frankish 2007, p. 657.

weaknesses are their conceivability premises and their appeal to CP. Since Frankish's anti-zombie conceivability premise is supposed to be roughly as plausible as Chalmers' *Conceivability Premise*, and for the same reasons, in so far as one conceivability premise should be considered implausible, they should both be considered implausible. Likewise, if CP is false, then both arguments employ false justification, as both employ CP. So, either way, Chalmers' argument is supposed to have a serious weakness.

Chalmers hasn't said much in response to the anti-zombie argument. He merely makes the following three points: (i) the anti-zombie conceivability premise is implausible; (ii) the hypothetical truth of Chalmers' *Conceivability Premise* and CP together support Chalmers' conclusion; and (iii) if Chalmers' *Conceivability Premise* and CP are a priori knowable, then the anti-zombie conceivability premise is not ideally conceivable (even negatively).[186] Thus Chalmers' response involves mere speculation and self-affirming hypotheticals, which suggests that he thinks the debate consists of a mere battle of intuitions.

Yet, if Chalmers thinks that the debate between he and Frankish consists of a mere battle of intuitions, I think he is wrong. For I will now give a straightforward critique of Frankish's argument that doesn't simply appeal to intuition.

The most obvious flaw is that line 2′ doesn't follow. For CP does not warrant inferences from ideal 1,2-conceivability to 1,2-possibility; it merely warrants inferences from ideal 1-conceivability to 1-possibility. Thus Frankish's parity of reasons strategy cannot work as intended. For he takes himself to employ the same CP principles as Chalmers, with the idea that, if either argument employs false justification by virtue of employing a CP principle, then both arguments employ false justification by virtue of employing that same CP principle. Yet we have seen that Frankish's argument is not even valid, as line 2′ does not follow even if CP is true.

---

[186] See Chalmers 2010, p. 180.

Thus, to make the anti-zombie argument valid, Frankish would have to employ a more powerful conceivability-to-possibility principle. For instance, CP can be combined with 2-CP:

      2-CP   Ideal 2-conceivability entails 2-possibility

CP and 2-CP together imply that ideal 1,2-conceivability entails 1,2-possibility; so CP and 2-CP can be used jointly to justify line 2′. This would make the anti-zombie argument valid.

Chalmers says in his paper, "Does Conceivability Entail Possibility?" that he thinks 2-CP is a plausible principle.[187] As he puts it, he thinks it is plausible that "Ideal secondary (positive/negative) conceivability entails secondary possibility."[188] However, he doesn't discuss it in much depth; he hardly says anything about it. However, what he does say about it is important. Namely, he says that ideal 2-conceivability is "deeply a posteriori [and thus] even if [it] is a guide to possibility, it will yield no a priori access to modality."[189] Thus the idea is that, although ideal 2-conceivability is plausibly a reliable guide to 2-possibility, ideal 2-conceivability judgments require empirical knowledge. This coincides with his remark about what it would take for a 2-conceivability judgment to be non-ideal and thus "go wrong as a guide to possibility." He writes:

> Prima facie secondary conceivability judgments can go wrong as a guide to secondary possibility when a subject is misinformed about relevant nonmodal empirical facts, and perhaps when an incautious subject is merely ignorant of those facts.[190]

The idea, once again, is that 2-conceivability judgments need to be properly informed about the right *empirical* facts in order to be a reliable guide to what is 2-possible. Thus, whereas ideal 1-conceivability judgments are supposed to be a priori, ideal 2-conceivability judgments will be, or will at least tend to be, a posteriori. The question arises, then, what standards would have to be

---

[187] Chalmers 2002a, § 6.
[188] Chalmers 2002a, § 6.
[189] Chalmers 2002a, § 4.
[190] Chalmers 2002a, § 6.

met in order for a 2-conceivability judgment to count as ideal. Presumably it would require that, in addition to involving no cognitive defect, one would have to know all of the relevant empirical truths. Given these hard-to-achieve standards, 2-CP is rendered plausible at the expense of a trade-off: it will often be difficult to make ideal 2-conceivability judgments in the first place. For this will often require a lot of hard-to-get empirical knowledge.

For instance, having an ideal 2-conceivability judgment about whether it is possible for the masked intruder to fly might be difficult to discern. At a bare minimum, it will require knowing more than the mere fact that the masked intruder is wearing a mask and is small enough to fit inside of the intruded upon space.

In the modified anti-zombie argument, the conceivability premise asserts the ideal conceivability that, corresponding to some possible being that has conscious experience, X, there is a 1,2-possible bare physical duplicate of this being that also has conscious experience X. Would an ideal conception of this require hard-to-get empirical knowledge? I argue that it would require very hard-to-get empirical knowledge. Indeed, since the actual world is the only world that one can get empirical knowledge about, knowing about the 1,2-possibility of a bare physical duplicate of a conscious being would require knowing that our world has a bare physical duplicate of a conscious being. But in that case, Frankish's conceivability premise could be made true only by someone who already knew that our world's consciousness is physical. This, I think, would be very hard to figure out by empirical investigation, especially as *ideal* 2-conceivability seems to require that the relevant empirical discoveries be known with certainty.

Thus, if Frankish were to modify his argument by employing both CP and 2-CP, this would have the benefit of rendering his argument valid, but it would have the drawback of

rendering his anti-zombie conceivability premise very implausible—much more implausible than Chalmers' *Conceivability Premise*.

Moreover, if Frankish's anti-zombie argument is going to work as a parity of reasons argument against Chalmers' argument, Frankish will have to put some constraints on what the intrinsic natures might be like. After all, if the intrinsic natures were phenomenal or protophenomenal, this would render the anti-zombie possibility a kind of Russellian monism, in which case Frankish's argument would not work as a parity of reasons argument against Chalmers' argument. For Frankish's strategy only works if his conclusion is inconsistent with Chalmers' conclusion. Yet Russellian monism is one of the disjuncts in Chalmers' conclusion. In his most rigorous formulation—i.e., his two-dimensional argument against materialism—his conclusion is that materialism is false or Russellian monism is true.

So, in order to work as a parity of reasons argument against Chalmers' argument, Frankish must require that the relevant token identities are between conscious events or properties and physical events or properties whose intrinsic natures are neither phenomenal nor protophenomenal. To be non-phenomenal, the intrinsic natures must not be conscious experiences. To be non-protophenomenal, the intrinsic natures (in their arrangement) must neither constitute nor play a key role in explaining our world's conscious events or properties. Making sure these constraints hold is going to make the anti-zombie conceivability premise even less plausible. For, in addition to having to know with certainty that our world's conscious events or properties are identical to various physical events or properties, we will also have to know that the physical intrinsic natures are neither phenomenal nor protophenomenal. Yet it is difficult to see how this could be known with certainty.

Altogether, one problem with Frankish's anti-zombie conceivability argument is that, unlike Chalmers' two-dimensional argument against materialism, it is invalid. For it appeals to CP principles, which are only supposed to work with 1-intensions, to get from 1,2-conceivability to 1,2-possibility. Yet, even when re-formulated so as to be valid, the anti-zombie argument's conceivability premise is still much less plausible than Chalmers' *Conceivability Premise*, as it requires *empirical certainty* that our world's consciousness is physical. Either way, the anti-zombie argument is not a good parity of reasons argument.

CHAPTER 6

PROBLEMS WITH DEFINING MATERIALISM

Chalmers' two-dimensional argument appeals to the conditional, $M_@$, which says that materialism is true of our world, $w_@$, only if $\Box(P_{1,2}T_1 \supset Q_1)$. $M_@$ thereby asserts a necessary condition of materialism's truth in $w_@$. Yet the question remains whether $M_@$ can be derived from a tenable definition of materialism. The purpose of this chapter is to investigate this. Its importance is that, if $M_@$ cannot be derived from a tenable definition of materialism, then Chalmers' two-dimensional argument appeals to a principle, $M_@$, whose only justification is that it is implied by an untenable definition.

I will begin by discussing options for how to define materialism that are *prima facie* tenable and that imply $M_@$. Then I will discuss how well these definitions fare against an objection from John Hawthorne (2002), an objection from Jaegwon Kim (1987/1993, 1989/1993, 1990/1993), and relevant Hempelian problems.

Ways of defining materialism

Ambiguous definitions of materialism (or "physicalism," as they sometimes say) are often given. It is particularly common to give formulations such as the following:

> Physicalism says that all facts, including all phenomenal facts, are metaphysically necessitated by the microphysical facts.[191]

This formulation is ambiguous in several ways. For instance, it is not clear what is supposed to be meant by a *fact*, or by *metaphysical necessitation*, or by *the* microphysical facts. However, let us now discuss some of the best options for how to define materialism more clearly.

---

[191] Kallestrup 2006, p. 273.

One of the best options is Frank Jackson's (1998/2000) definition of materialism. He says that materialism is true of a world *w* iff every minimal microphysical duplicate is a duplicate simpliciter.[192] Here "minimal" serves a "that's all" function, and "duplicate simpliciter" implies a "that's all" function. Jackson describes "that's all" clauses as "stop" clauses:

> A minimal [micro]physical duplicate is what you get if you 'stop right there'. (Writers of recipes and construction manuals typically rely on an intuitive understanding of an implicitly included 'stop' clause in their recipes; otherwise they would face the impossible task of listing all the things you should *not* do.) Thus, a minimal physical duplicate of our world is a world that (a) is exactly like our world in every physical respect (instantiated property for instantiated property, law for law, relation for relation), and (b) contains nothing else in the sense of nothing more by way of kinds or particulars than it *must* to satisfy (a). Clause (b) is a 'no gratuitous additions' or 'stop' clause.[193]

The purpose of the first "that's all" clause can be demonstrated with a food analogy. If I say that I have a good recipe that I'd like you to try, I won't make any guarantees for how it will turn out if one adds extra ingredients to the mix. I might claim, for instance, that my tomato soup doesn't have much bite; however, I won't guarantee that it doesn't have much bite when mixed with Vodka and hot sauce. Likewise, if someone prepares the recipe as specified with no extra ingredients or steps, there should be no difference in outcome, taste-wise. The food should taste the same—not merely, say, at least as salty. Thus I might claim, of my dish, that a minimal duplicate of its ingredients and steps of preparation will result in a duplicate simpliciter. The idea is that, as long as the ingredients and preparation are the same, the outcome will be the same.

In a similar way, Jackson (2000) says that materialism is true of world *w* iff any minimal microphysical duplicate is a duplicate simpliciter. Here, as always, "micro" means basic/fundamental, with no connotation of small size. It may be that all fundamental properties are small, but we don't need to assume this. The idea, then, is that materialism is true of *w* iff any possible world that has the same positive microphysical truths (and no other basic positive truths)

---

[192] Jackson 2000, p. 12.
[193] Jackson 2000, p. 12-13.

as *w* will have the same sum of all positive truths as *w*, thereby duplicating *w* in every qualitative respect.[194] We can state Jackson's definition more formally as follows:

> D$_{jac}$ Materialism is true of *w* iff □[(*w*'s positive microphysical truths and that's all the positive micro truths) ⊃ (*w*'s positive truths and that's all the positive truths)].

Recall: the present goal is to investigate whether M$_@$ follows from a tenable definition of materialism. We will now see that D$_{jac}$, which is prima facie tenable, implies M$_@$.

D$_{jac}$ implies M$_@$. For, in D$_{jac}$, if we plug in our world, $w_@$, for *w* and introduce "T‴" to stand for "that's all the positive truths," we get: Materialism is true of $w_@$ iff □(PT⊃AT′). But according to Chalmers' two-dimensional semantics, 1- and 2-intensions are automatically at play when the speaker doesn't specify otherwise. Hence D$_{jac}$ implies that materialism is true of $w_@$ iff □(P$_{1,2}$T$_{1,2}$⊃A$_{1,2}$T′$_{1,2}$), hence iff □(P$_{1,2}$T$_1$⊃A$_{1,2}$T′$_{1,2}$), hence only if □(P$_{1,2}$T$_1$⊃Q$_1$).

Another popular way of defining materialism is Chalmers' (1996) definition. He says that materialism is true of *w* iff any microphysically indiscernible world has at least the positive truths of *w*.[195] Let us re-state it more formally as follows:

> D$_{cha}$ Materialism is true of *w* iff □[(*w*'s positive microphysical truths and that's all the positive microphysical truths) ⊃ (*w*'s positive truths)].

By contrast with D$_{jac}$, D$_{cha}$ only gives one "that's all" clause, and it's a different one. Whereas D$_{jac}$ says that's all the positive *micro* truths (microphysical or micro-nonphysical), D$_{cha}$ says that's all the positive *microphysical* truths. *Ceteris paribus*, this makes materialism *prima facie* less likely to be true on D$_{cha}$. For the difference is that, on D$_{cha}$, materialism's truth requires that our world's mental properties be instantiated in any microphysically indiscernible world,

---

[194] It is important to say every *qualitative* respect. For it must instantiate the same properties, but it doesn't matter which individuals instantiate them.
[195] See Chalmers 1996, p. 41; although note that earlier on p. 33 he says, "talk of physical properties is implicitly restricted to the class of fundamental properties unless otherwise indicated. I will sometimes speak of "microphysical" or "low-level physical" properties to be explicit."

regardless of which micro-nonphysical truths obtain. Yet for all we know prima facie, there might be some possible micro-nonphysical truths that would prevent some of our world's mental properties from being instantiated. Hawthorne (2002) emphasizes this, as we will see.

Nevertheless D_{cha} is at least *prima facie* tenable, and it implies $M_@$. We can show that D_{cha} implies $M_@$ in the following way: in D_{cha}, if we plug in $w_@$ for w and introduce "T*" to stand for "that's all the positive *microphysical* truths," we get: Materialism is true of $w_@$ iff $\Box(PT^* \supset A)$, hence iff $\Box(PT \supset A)$, hence iff $\Box(P_{1,2}T_{1,2} \supset A_{1,2})$, hence iff $\Box(P_{1,2}T_1 \supset A_{1,2})$, hence only if $\Box(P_{1,2}T_1 \supset Q_1)$.

Next I will discuss John Hawthorne's (2002) blocker arguments against each of these *prima facie* tenable definitions: D_{jac} and D_{cha}.

Hawthorne's "blocker" arguments

John Hawthorne (2002) gives a "blocker" argument against Jackson's definition of materialism, D_{jac}, and Chalmers' definition of materialism, D_{cha}.

A key piece of terminology here is that of a *blocker*. As discussed in chapter 5 (when we discussed Leuenberger's blocker argument against Chalmers' *Conceivability Premise*), a blocker would be a property whose instantiation prevents the instantiation of certain mental properties.[196] Hawthorne is particularly interested in the possibility of non-physical blockers. A non-physical blocker would be a non-physical property whose instantiation prevents the instantiation of certain mental properties. Hawthorne confesses that he doesn't know whether non-physical blockers are possible; he is agnostic on this matter. However, he argues that we should not define materialism in such a way that assumes that they are impossible. For as he says: "when defining

---

[196] Strictly, a blocker would be a *property or being* whose existence or instantiation can prevent the instantiation of a conscious property, but I'll tend to speak in terms of blockers as properties for simplicity.

materialism, one should not take a stand on controversial modal issues about which, intuitively, the materialist qua materialist has no commitments."[197]

Hawthorne also holds that a suitable definition of materialism should imply a strong physical dependence principle. Although he does not formulate such a principle, he makes clear that he at least accepts the following:

PD   Materialism is true only if every truth (positive or negative) depends solely on physical truths (positive or negative), hence only if every (instantiated or non-instantiated) mental property depends solely on (instantiated or non-instantiated) physical properties.[198, 199]

According to PD, materialism is false if there is a positive mental truth that obtains partly because of the non-instantiation of a non-physical blocker. For in such a case the positive mental truth would not depend solely on physical truths, but would depend at least partly on the non-instantiation of a non-physical blocker. (Hawthorne implicitly holds that the *non-instantiation* of a *non-physical* blocker would not count as a physical truth.) Hawthorne judges that this is a good result, and that any proper definition of materialism should imply this.

Yet as Hawthorne points out, D_{jac} does not imply this. For, according to D_{jac}, the only way materialism can be false is if there is a minimal physical duplicate that is not a duplicate simpliciter. Thus, according to D_{jac}, if a positive conscious truth obtains partly because of the absence of certain non-physical blockers, this would not constitute a counterexample to materialism, as it would not be a case of a minimal physical duplicate failing to be a duplicate simpliciter. Indeed, a world with non-physical blockers cannot be a *minimal* physical duplicate

---

[197] Hawthorne 2002, p. 110.
[198] See Hawthorne 2002, p. 109.
[199] Hawthorne's rationale for including both positive and negative truths is as follows: "Materialists about mentality will not tolerate the existence of mental facts, positive or negative, that have a robust explanation that appeals to immaterial entities" (Hawthorne 2002, p. 109).

of anything, let alone a minimal physical duplicate that is not a duplicate simpliciter! After all, by the definition of *minimal*, a minimal physical duplicate cannot have any non-physical beings or properties, and thus cannot have any *non-physical* blockers. So, since worlds that have non-physical blockers are not minimal physical duplicates of anything, they cannot be minimal physical duplicates of our world, and therefore they cannot be minimal physical duplicates of our world that fail to be duplicates simpliciter of our world.

So, as Hawthorne puts it, D_{jac} might be able to correctly assess whether a world's physics "weakly necessitate[s]"[200] its positive conscious truths in the sense of necessitating them in all of the worlds that don't have blockers; but it won't be able to correctly assess whether it "metaphysically necessitate[s]"[201] the positive conscious truths in the sense of necessitating them in all of the worlds including the blocker worlds. Indeed, D_{jac} is problematic, since (for all we know) there might be a world w whereby D_{jac} mistakenly judges materialism to be true of w even though PD correctly judges materialism to be false of w.

By contrast Chalmers' (1996) definition, D_{cha}, effectively defines the realm of counterexamples as the set of all worlds that are microphysically indiscernible from w. In this way, worlds that have non-physical blockers *can* function as counterexamples to materialism's truth in w, since worlds that have non-physical blockers can be microphysically indiscernible from w. Hawthorne is happy with D_{cha} in this regard.

However, Hawthorne emphasizes that D_{cha} is like most definitions of materialism in that it only gives a definition of what it would take for materialism to be true of a *world*; yet he says that many philosophers also want a definition of what it would take for materialism to be true of a *proper part* of a world, or some part of a world besides the total world. As he puts it,

[200] See Hawthorne 2002, p. 103 and p. 111 fn. 1.
[201] See Hawthorne 2002, p. 103 and p. 111 fn. 1.

"Philosophers are interested not merely in providing an account of materialism *simpliciter*, but also an account of materialism about a certain class of properties P (mental properties, for example) – call these 'restricted materialisms'."[202]

Hawthorne's gripe with D_{cha} then becomes the following: if D_{cha} were to be suitably modified to say what it would take for materialism to be true of some restricted class of properties, the modified version would have a blocker problem just as serious as D_{jac}'s blocker problem. For the way that Chalmers would modify D_{cha} to test for "restricted materialisms," Hawthorne thinks, is clear. He says: "It is clear how Chalmers thinks such a definition would go. Materialism about, say, the mind, is true just in case every world physically indiscernible from the actual worlds [sic] will have at least the positive mental facts that obtain at the actual world."[203] Thus we get a proposal for Chalmers' definition of materialism when restricted to cover just the mental truths:

> $_rD_{cha}$ Materialism is true of the mental truths iff every physically indiscernible world has at least the positive mental facts that obtain in our world.

Likewise, a version of PD (restricted to the mental truths) would go as follows:

> $_rPD$ Materialism is true of the mental truths only if the mental truths (positive or negative) depend solely on physical truths (positive or negative), hence only if every (instantiated or non-instantiated) mental property depends solely on (instantiated or non-instantiated) physical properties.
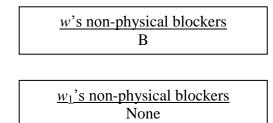
But now Hawthorne says that there is a problem. For suppose there's a non-physical blocker that blocks the instantiation of some mental truth. In that case, $_rPD$ will rightly judge that materialism is false of the mental truths, since there's a negative mental truth that does not depend solely on

---

[202] Hawthorne 2002, p. 108.
[203] Hawthorne 2002, p. 108. Note: for our purposes we can ignore that this formulation is only about "the mind" and not general to allow any kind of property to be plugged in.

positive or negative physical truths, but depends partly on the instantiation of the non-physical blocker. Yet, as we will see, $_rD_{cha}$ is not guaranteed to yield the same result.

For instance, suppose world w has some positive mental truths. Now consider the set of all worlds that are microphysically indiscernible from w: say, $\{w, w_1\}$. Suppose that, of these worlds, w has exactly one non-physical blocker, B, which blocks some mental property, whereas w_1 does not have any non-physical blockers. That is, suppose the distribution is as follows:

w's non-physical blockers
B

w₁'s non-physical blockers
None

Is materialism true of w's mental properties? As for this question, $_rPD$ and $_rD_{cha}$ will disagree. $_rPD$ will judge materialism to be **false** of w's mental properties, since w has a negative mental truth that depends partly on the instantiation of its non-physical blocker, B. Meanwhile $_rD_{cha}$ will judge materialism to be **true** of w's mental properties, since all of the worlds that are physical indiscernible from w (i.e., w and w_1) have *at least* the mental properties of w. So, given the possibility of non-physical blockers, $_rD_{cha}$ and $_rPD$ can yield contradictory results. Hawthorne takes this as a reason to reject $_rD_{cha}$. Indeed, he goes further and takes this problem for $_rD_{cha}$ as an indirect problem for D_{cha} on the grounds that $_rD_{cha}$ is how D_{cha} would be "suitably modified" to define materialism of the mental.

In sum, there are two non-physical blocker cases that Hawthorne is worried about. The first is a case where a positive mental truth obtains partly because of the non-instantiation of some non-physical blocker. This is the option that threatens D_{jac}. In such a case, PD implies that materialism is false, since there is an instantiated mental property that depends partly on the non-

instantiation of some non-physical blocker. The second is a case where a negative mental truth obtains partly because of the instantiation of a non-physical blocker. This is the option that threatens $_rD_{cha}$. In such a case, $_rPD$ implies that materialism is false, since there is a non-instantiated mental property whose non-instantiation depends partly on the instantiation of some non-physical blocker.

According to Daniel Stoljar's (2009) *Stanford Encyclopedia of Philosophy* entry for "Physicalism" (edited by David Chalmers), "[Hawthorne's] blockers problem . . . is an open question in the literature."

I am suspicious of Hawthorne's blocker arguments. Indeed, I argue that non-physical blockers are impossible, and thus I argue that materialism should not have to be defined in such a way as to allow that non-physical blockers can serve as counterexamples to materialism.

I argue that non-physical blockers are impossible. For the true physics at least catalogs structure, and presumably it will catalog any structure regardless of what kind of intrinsic natures are involved. Thus anything that contributes structure will count as physical. But it is plausible that blockers would contribute structure. For presumably it makes a structural difference whether a given blocker exists. For instance, 10^{17} entities plus zero blockers should equal 10^{17} entities, whereas 10^{17} entities plus 28 blockers should equal $10^{17} + 28$ entities. So, since it is plausible that blockers would contribute structure, and since it is plausible that contributing structure entails being physical regardless of what one's intrinsic natures are like, it is therefore plausible that blockers would have to be physical, in which case non-physical blockers are impossible. Yet, even if blockers do not contribute structure, they still seem impossible. For, if they do not contribute structure, they must have roughly the existential status of numbers. For that seems to be the only way of existing without making a structural contribution. Yet, on this way of

existing, non-physical blockers would be abstract. But being abstract should imply causal impotency, in which case they should not be able to block anything, and thus should not be able to function as blockers. Altogether, then, non-physical blockers either contribute structure or they do not. If they do, they are physical. If they do not, they are abstract like numbers. But if they are abstract like numbers, they are causally impotent, and thus cannot block anything—or at least we cannot understand how they could block anything. On the former horn, non-physical blockers would fail to be non-physical. On the latter horn, non-physical blockers would fail to be blockers. Either way is contradictory. Thus non-physical blockers are impossible.

Assuming non-physical blockers are impossible, the only remaining question is whether materialism can be defined in a tenable way without incorporating PD. It is plausible, after all, that materialism should imply some sort of physical dependence principle; the only question is how strong a physical dependence principle should be. If non-physical blockers are impossible, PD will be overkill provided a sufficiently strong alternative physical dependence principle is in the offing. I wish now, then, to propose an alternative physical dependence principle that doesn't imply PD. Namely, I suggest that the following is strong enough:

PD*   Materialism is true only if every instantiated micro property is microphysical. PD* captures the (physical dependence) intuition that materialistic worlds cannot have truths that depend on the instantiation of micro-*nonphysical* properties. I doubt a definition of materialism should need to imply a stronger physical dependence principle than this.

D_{jac} implies PD*. For PD* is equivalent to the thesis that materialism is false of any world where there is at least one instantiated micro-nonphysical property. Yet D_{jac} implies this. For such a world would not be a duplicate simpliciter of any minimal physical duplicate, as it would have at least one instantiated micro-nonphysical property that no minimal physical

117

duplicate would have. Thus D_{jac} implies PD*, which is a tenable physical dependence principle that is *prima facie* strong enough to capture materialist intuitions about physical dependence. The only remaining question is whether PD* is strong enough to capture *all* of the right physical dependence intuitions about materialism. However, given the plausibility that non-physical blockers are impossible, it is difficult to see what other physical dependence intuitions should be captured. Thus it seems that PD* is sufficient and therefore D_{jac} implies all the right physical dependence principles. Thus D_{jac} seems safe from Hawthorne's blocker critique.

What about Hawthorne's critique of D_{cha}? This critique also requires taking the possibility of non-physical blockers seriously. Yet I argued that non-physical blockers are impossible. So, the only remaining question is whether D_{cha} and $_rD_{cha}$ imply sufficiently strong physical dependence principles.

D_{cha} implies PD*, which we've seen to be at least a *prima facie* sufficiently strong physical dependence principle. We can prove that D_{cha} implies PD* as follows. PD* says that materialism is false of any world where there is at least one instantiated micro-nonphysical property. Yet D_{cha} implies this. For relative to any world w that has some instantiated micro-nonphysical property, φ, there will be at least one world that is microphysically indiscernible from w that does not have all of w's positive truths: namely, a minimal microphysical duplicate of w won't have φ, since if it did then it wouldn't be a *minimal* microphysical duplicate.

Nevertheless, $_rD_{cha}$ does not imply PD*, since it restricts its claim to the mental facts instead of to all of the world's facts. However, $_rD_{cha}$ does imply a restricted version of PD* that restricts its claim to the instantiated micro properties that are relevant to the instantiated mental properties: namely,

$_r$PD*   Materialism is true of the mental only if every instantiated micro property (upon

which some instantiated mental property depends (wholly or partly)) is

microphysical.

It does not seem that $_r$D$_{cha}$ should need a physical dependence principle that is any stronger than

$_r$PD*, and yet we can prove that $_r$D$_{cha}$ implies $_r$PD* as follows. $_r$PD* says that materialism is false

of the mental in any world where there is at least one instantiated micro-nonphysical property

upon which some instantiated mental property depends. Yet $_r$D$_{cha}$ implies this. For relative to any

world w that has some instantiated micro-nonphysical property upon which some instantiated

mental property, φ, depends, there will be at least one world that is microphysically indiscernible

from w but that does not have all of w's positive mental truths. Namely, a minimal microphysical

duplicate of w would be microphysically indiscernible but wouldn't have φ, since if it did then it

wouldn't be a *minimal* microphysical duplicate. Thus a minimal microphysical duplicate of w

wouldn't have w's positive mental truths that depend on φ, since it wouldn't have φ. Altogether,

then, $_r$PD* merely gives a necessary condition of materialism of the mental, but since any

counterexample to $_r$PD* would also be a counterexample to $_r$D$_{cha}$, $_r$PD* is a necessary condition

of $_r$D$_{cha}$, i.e., $_r$D$_{cha}$ implies $_r$PD*.

There is a further consideration to strengthen my rebuttal to Hawthorne's blocker

problem for $_r$D$_{cha}$. Namely, that, even if one is interested in restricted materialisms, it is not clear

that this requires giving a *restricted definition* of materialism. On the one hand, perhaps some

conglomerations of some classes of properties can be instantiated in worlds where they are the

only instantiated properties in the world. That is, perhaps there can be a world consisting of

precisely that conglomeration of instantiated properties and nothing else. In such a case, since

D$_{cha}$ defines materialism of a *world*, it thereby defines materialism of that restricted class of

properties, since it defines materialism of a world consisting of just that conglomeration of properties. On the other hand, even if some conglomerations of properties cannot be instantiated in worlds that have no other instantiated properties, it is not clear that such conglomerations deserve a restricted definition of materialism that gives conditions for whether materialism is true of them alone. For in so far as they cannot exist in a world all by themselves, they are not self-subsisting but instead are connected in some way to other properties. Thus, if we are going to ask whether materialism is true of them, presumably we should be asking instead whether materialism is true of a minimal conglomeration that they are connected with. But in so far as a minimal conglomeration is defined partly as needing to be able to be instantiated all by itself in a world of its own, then once again we don't need a restricted definition of materialism to cover such cases; all we need is a definition of materialism of a world.

Altogether, D_{cha} implies a physical dependence principle, PD*, that seems strong enough to capture appropriate physical dependence intuitions with respect to materialism's truth of a world. Meanwhile, although I just argued that restricted definitions of materialism are unnecessary since any minimal conglomeration of properties could exist in a world of its own anyway and therefore be covered by a definition of materialism of a world, I also argued that, even if restricted definitions of materialism are necessary, $_rD_{cha}$ implies a physical dependence principle, $_rPD*$, that seems strong enough to capture appropriate physical dependence intuitions with respect to materialism's truth of the mental.

In sum, we have seen that Hawthorne's blocker arguments are serious problems for Jackson's (2000) definition of materialism, D_{jac}, and Chalmers' (1996) definition of materialism, D_{cha}. Yet I argued that non-physical blockers are impossible, which renders the blocker arguments naïve. I also pointed out that D_{jac} and D_{cha} both imply the physical dependence

principle, PD*, which seems to be a sensible replacement for PD; and that $_rD_{cha}$ implies the restricted physical dependence principle, $_rPD^*$, which seems to be a sensible replacement for $_rPD$. Finally, I argued that, even if one is interested in restricted materialisms, this shouldn't require giving a restricted definition of materialism, in which case Hawthorne is wrong to think that D_{cha} would have to be "suitably modified" to define restricted materialisms.

Next I will discuss a potential problem from Jaegwon Kim.

Kim's "lone ammonium molecule" argument

Jaegwon Kim's (1987/1993, 1989/1993, 1990/1993) "lone ammonium molecule" problem is one of the most important problems for defining materialism. Kim's intuition is that materialism should be true only if there are what he calls "local determinations" between various types of brain states (say, C-fiber excitation), and various types of mental states (say, pain). A local determination between C-fiber excitation and pain would be a necessary connection between the two, so that C-fiber excitation entails pain in the sense that any individual that exhibits C-fiber excitation also experiences pain. So that, necessarily, if there is C-fiber excitation, then there is pain regardless of what the rest of the world is like. Again, this is an example of what Kim would call a "local determination." The rationale for calling it "local" rather than "global" is that each object determines its own mental properties regardless of what the rest of the universe (or "globe") is like.

Kim thinks that, if materialism is true of the actual world $w_@$, then there should be many local determinations. Unfortunately, D_{jac} and D_{cha} are written in such a way that they don't test for specific local determinations; they only test for global determinations. They only test to see whether every truth in the world supervenes in the right way on every physical truth in the world. Thus D_{jac} and D_{cha} are subject to Kim's lone ammonium molecule objection:

It is consistent with [such a] version of materialism for there to be a world which differs physically from this world in some most trifling respect (say, Saturn's rings in that world contain one more ammonia molecule) but which is entirely devoid of consciousness, or has a radically different, perhaps irregular, distribution of mental characteristics over its inhabitants (say, creatures with brains have no mentality while rocks are conscious). As long as that world differs from this one in some physical respect, however miniscule or seemingly irrelevant, it could be as different as you please in any psychological respect you choose.[204]

The idea here is that, if a definition of materialism doesn't require specific local determinations, it is apt to allow that miniscule physical differences can make big mental differences. Yet elsewhere Kim points out another kind of absurdity. Namely, if specific local determinations are not required in order for materialism to be true, then materialism can be true, even if "within a given world, perhaps this one, [there might be] two physically indistinguishable organisms with radically different psychological attributes."[205] The idea here is that, within a single world, multiple people might be physically indiscernible, yet mentally discernible in some radical way. For, if materialism does not require entailments between specific brain state types and specific mental state types, materialism can be formally consistent with the option that different people in the same world can have the same brain states yet have radically different mental states.

The best response in the literature comes from Robert Stalnaker. For starters, he agrees with Kim that it is not "sensible" to accept the possibility of "lone ammonium molecule" type cases. He writes: "I agree that no *sensible* materialism would accept the possibility of a world that differed physically from ours only in this way, while being radically different from it in the distribution of mental properties."[206] Yet he disagrees with Kim that this is a problem for definitions of materialism that don't take specific local determinations as necessary conditions of materialism's truth. For he argues: "One should not define materialism so that there cannot be

---

[204] Kim 1987/1993, p. 85-6; cf. Kim 1989/1993, p. 277; Kim 1990/1993, p. 155.
[205] Kim 1987/1993, p. 86; cf. Kim 1989/1993, p. 277-8; Kim 1990/1993, p. 155.
[206] Stalnaker 1996, p. 229, *emphasis added*.

122

silly versions of it."[207] Thus Stalnaker argues that, although it would be silly to believe in "lone ammonium molecule" type cases, one should not define materialism in a way that would render materialism false if such cases are possible.

Stalnaker defends his view by analogy with a hypothetical example of a philosopher who holds the following (silly) view:

> The mental properties of human beings on earth [are] *constituted*, in part, by the precise distribution of hydrogen atoms in distant parts of space.[208]

The apparent absurdity of this view is even more severe than the apparent absurdity of the lone ammonium molecule type cases; but, although Stalnaker goes on to say that the view would be "bizarre and unmotivated,"[209] he says that it is nevertheless consistent with *materialism*, since it is consistent with the basic materialist idea that "everything in the world is physical."[210]

Furthermore, since it is not a generally accepted principle of theoretical definitions that the theories defined can only be believed in sensible ways, the burden should be on Kim to make an exception for the definition of materialism. Why should we require that materialism can only be accepted in sensible ways? Kim doesn't give such an argument; nor can I think of a good one. By contrast, Stalnaker's proposal is plausible: namely, that the essence of materialism (in most philosophers' minds, anyway) is the idea that everything in the world is physical. As such, this is all that a definition of materialism should capture; it should not incorporate further constraints just for the sake of minimizing the number of people with bizarre physical theories from counting as materialists. Likewise, a person shouldn't be prevented from counting as a materialist just for having, say, a low IQ.

---

[207] Stalnaker 1996, p. 229.
[208] Stalnaker 1996, p. 229-30.
[209] Stalnaker 1996, p. 230.
[210] Stalnaker 1996, p. 230.

If Kim wants to make an exception for the definition of materialism so as to rule out the possibility of "lone ammonium molecule" type cases, he should have to argue that a definition of materialism should be trying to capture something along the lines of the following:

Materialism is true iff both (i) everything is physical, and (ii) digestion, table surfaces, conscious experiences, etc., are determined by *local* physical properties.

I doubt this is the basic intuition that materialists generally wish to capture in a definition of materialism, as I doubt that clause (ii) is commonly seen as crucial. Rather, quibbling about which properties (if any) are locally determined seems to be (in most philosophers' minds) a separate issue. For instance, a person might think materialism is true, but remember hearing something about Bell's theorem as evidence for non-locality of causation. Or a person might think materialism is true, but believe that causation of consciousness is sometimes a non-local phenomenon. In either case, it seems that it shouldn't be the job of a definition of materialism to take a stand on what should rather be a footnote about our world's physics.

Finally, it is important to point out that, even if such a definition as Kim would prefer were to be constructed, $M_@$ would follow from it anyway. For suppose a definition of materialism were to be constructed that has all of the features that Kim desires. This would be a definition that takes many specific local determinations as necessary conditions of materialism's truth *in addition* to taking global determinations as necessary conditions of materialism's truth. For the problem with global determinations, in Kim's mind, is not that they don't exist, but that they don't fully capture the materialist's intended commitments. Thus on a definition of materialism that Kim would endorse, materialism's truth would imply the determinations of the global microphysical truths to the global truths; in that case it would imply $M_@$, since one of the global determinations in our world would be that $P_{1,2}T_1 \supset Q_1$. Thus, from any definition of

materialism that Kim would endorse, $M_@$ would follow as a consequence.[211] This is all Chalmers

needs for the purpose of his two-dimensional argument, as he merely needs it to imply $M_@$ so

that he can cite $M_@$ in a modus tollens operation.

Hempelian problems

The objections from Hawthorne and Kim are objections to particular definitions of materialism.

Now we will discuss a different class of problems: Hempelian problems. Hempelian problems

are named after Carl Hempel. The basic problem is supposed to be that, given the way

materialism is defined, the truth of materialism is either trivial (trivially true or trivially false) or

else unknowable by a priori means. Either way, philosophizing about whether materialism is true

is unlikely to be worthy of a great deal of one's time.

I am going to proceed with some historical context, as Hempel himself couches his

discussion mainly in terms of whether biology is reducible to physics and chemistry. He says that

his points generalize to similar problems (such as the materialist debate), but most of his

examples are regarding biological reductionism. So what I am going to do is to explain his points

regarding biological reductionism, and then apply them to the materialism debate.

Hempel says that there are two ways that we can phrase the question of biological

reductionism. One way is to phrase the question in terms of whether the *theories* of biology are

reducible to the *theories* of physics and chemistry. The other way is to phrase the question in

terms of whether our world's *actual* biology is reducible to our world's *actual* physics and

chemistry. He argues that either way there are problems.

On the one hand, suppose we ask whether the theories of biology are reducible to the

theories of physics and chemistry. Hempel points out that, on this construal, the theories of

biology will include all of the theories of biology that have ever been given, including "theories

---

[211] $M_@$ would not be a competing definition; indeed it is not a definition at all, it is merely an "only if" statement.

125

that have been discarded as empirically false or as untenable on other grounds."[212] Yet he points out that this way of phrasing the question doesn't capture the "philosophical intent"[213] of the question.[214] That is, it is not the question that philosophers intend to be asking. They don't mean to be asking whether some biological theories (that for all we know are false) are reducible to some physics and chemistry theories (that for all we know are false). The intended question is not about theories that have been given; it is about the nature of our world.

So, Hempel believes that the better way to phrase the question of whether of biology is reducible to physics and chemistry is to ask whether our world's *actual* biology is reducible to our world's *actual* physics and chemistry. As he puts it, the thesis that biology is reducible to physics and chemistry should be cashed out in such a way that "[biology] refers only to 'actual' aspects of biological phenomena."[215] Equivalently, we can phrase it in terms of the *true* biological theories; that is, in terms of whether "Every true biological theory B is reducible to some true physicochemical theory P,"[216] as this way of phrasing the question captures the philosophical intent of the question.

Unfortunately, this latter way of phrasing the question has problems of its own. For, as Hempel points out, "the phrase 'every true biological theory' must be understood to refer . . . to any . . . true biological theory, whether or not it is actually ever thought of and formulated."[217] The problem here is that, although we are discussing the actual biology, we haven't yet discovered the actual biology: that's why biologists are still studying biology. Thus the actual

---

[212] Hempel 1969, p. 182.
[213] Hempel 1969, p. 183.
[214] E.g., Hempel says the reductionist thesis "cannot be saddled with the view that expressions such as 'vital force' or 'animal magnetism' are definable in physicochemical terms" (Hempel 1969, p. 182).
[215] Hempel 1969, p. 182.
[216] Hempel 1969, p. 182.
[217] Hempel 1969, p. 182.

biology is, as Hempel puts it, "an extremely elusive idea."[218] The elusiveness comes from the

fact that we don't yet know all of the true biological theories, let alone all of the true theories of

chemistry and physics; so how are we supposed to make arm-chair progress on the question of

whether the true biological theories are reducible to the true theories of chemistry and physics?

On the contrary, it appears that, since we don't yet know the true theories, we don't yet fully

understand the question; and we won't fully understand the question until we discover the true

theories via empirical research. Thus, as of now, there is simply no way that *arm-chair* study of

the question is going to be able to figure out the answer.

To make matters worse, the true theories might not be neatly dividable into the categories

of "physicochemical" and "biological" anyway.[219] As Hempel puts it:

> In the course of future research, the dividing line between biology and physics-and-chemistry may become as blurred as that between physics and chemistry has become in our time. Future theories might well be couched in novel kinds of terms functioning in comprehensive theories that afford explanations both for phenomena now called biological and for others now called physical or chemical. To the vocabulary of such a comprehensive unifying theory, the division into physico-chemical terms and biological terms might no longer be significantly applicable, and the notion of eventually reducing biology to physics and chemistry would lose its meaning.[220]

The idea here is that the *true* nature of our world might be such that the categories in question—

physics, chemistry, biology—are "[not] significantly applicable." For there might be

fundamental properties that we do not currently speak about but which "afford explanations both

for phenomena now called biological and for others now called physical or chemical."

Ultimately, then, the worry is that we might need a revisionist science to capture the way that our

world is; yet on a revisionist science, physics, chemistry, and biology—as we now understand

them—might not be real categories, in which case the question of whether biology is reducible to

---

[218] Hempel 1969, p. 182.
[219] See Hempel 1969, p. 182-3.
[220] Hempel 1966, p. 106.

physics and chemistry would be a naive question, presupposing categories that don't apply neatly to our world.

In sum, the question of whether biology is reducible to physics and chemistry can either be expressed as the question of whether biological *theories* are reducible to physicochemical *theories*, in which case the question lacks the philosophical intent of the people who ask it, since it includes theories that for all we know might be false; otherwise the question can be phrased in terms of the *true* physics, chemistry, and biology, in which case there is either an answer to the question but it is epistemically inaccessible to armchair reasoning, or else there is no answer to the question since the categories of physics, chemistry, and biology—as we now understand them—don't neatly apply to our world.

Hempel's dilemma does not only apply to biological reductionism. A dilemma of this kind, he says, "besets all reductionist theses that are conceived as ontological claims."[221] For any reductionist thesis can be split up into two separate questions: first, whether a given batch of *expressed* theories can be reduced to another given batch of *expressed* theories; and second, whether a given batch of *true* theories can be reduced to another given batch of *true* theories. The former is liable to lack the philosophical intent of the question, while the latter is liable to be naïve in virtue of presupposing categories that don't neatly apply to the world. We will see next that the same sort of dilemma applies to the traditional materialist debate.

Are conscious experiences reducible to physics? This question can be articulated in two different ways. The "*Currentist*" way is to ask whether our conscious experiences are reducible to our current physics theories, which are theories that we have already expressed.[222] The

---

[221] Hempel 1969, p. 183.
[222] I borrow the "Currentist" terminology from Crook and Gillett (2001).

128

"*Futurist*" way is to ask whether our conscious experiences are reducible to the *actual* physics, which might be discoverable—if at all—in the future.[223] Either way is problematic, as we'll see.

On a *Currentist* articulation, materialism will be true only if all of our world's positive truths are properly grounded in certain claims that we currently believe about our world's physics. Yet some positive truths are not grounded in what we already believe about our world's physics: for instance, the number of electrons on Pluto is not entailed by what we already believe about our world's physics. So, on a *Currentist* definition of materialism, materialism is false if for no other reason than that we don't currently believe anything that implies how many electrons are on Pluto. Thus the *Currentist* articulation doesn't do justice to the philosophical intent of the question of whether materialism is true. When we ask whether materialism is true, we don't mean to be asking, even partly, whether we already believe things that imply all of the facts in the world, including how many electrons are on Pluto.

Since *Currentism* is untenable, nearly everyone in the literature is a *Futurist*.[224] E.g., David Armstrong cashes out the relevant physics of "materialism" in terms of "whatever set of properties the physicist in the end will appeal to."[225] Barry Loewer cashes it out in terms of "the true comprehensive fundamental physical theory."[226] Frank Jackson cashes it out in terms of a "*completed* physics"[227] (the idea of "completed physics" being that, someday, physics might get completed in the sense that we will then have discovered the true physics). Michael Philips

---

[223] I borrow the "Futurist" terminology from Crook and Gillett (2001).
[224] The only exception is Melnyck (2003).
[225] Armstrong 1991, p. 186.
[226] Loewer 1996, p. 103. Loewer's formulation is my favorite, since it doesn't turn on whether the true physics is discoverable in principle. Yet for this very reason, it is the least deserving of the name "Futurist." But for convenience's sake, let us use the "Futurist" terminology anyway as a loose way of focusing on the true physics.
[227] Jackson 1986, p. 291.

(2013) cashes it out in terms of "physics' final words (if any)."[228] Chalmers, too, understands the materialist thesis as having to do with the true or "final" physics.

Unfortunately, the *Futurist* option is one according to which the materialist debate is problematic. One problem is that, since we don't know how the true physics goes, it should be difficult to have a worthwhile arm-chair discussion about whether our conscious experiences are reducible to, or entailed by, the true physics. At a conceptual level, if we don't know how the true physics goes, we don't fully understand the question of whether materialism is true, as we don't quite understand what it is about which we're wondering when we wonder whether our conscious experiences are reducible to, or entailed by, the true physics. As Philips puts it, "We could say that matter is whatever physicists finally decide it is. But this reduces materialism to a blank check to be filled in when physics finally closes its book."[229]

Indeed, as Noam Chomsky (1968), Tim Crane and Hugh Mellor (1990), Barbara Montero (1999), Torin Alter and Robert Howell (2009) and others have pointed out, prior to knowing more about the true physics, we cannot rule out the option that the true physics will consider "the mental" as fundamental—for instance, that panpsychism is true. As Montero puts it,

> Since we cannot predict the course of physics, we cannot even say with certainty that a final physics will not include mental properties, qua mental, as a [sic] fundamental properties. Yet if this purported final physics takes the mental realm to be fundamental, the significant difference between physicalists who claim that mental properties will be accounted for in terms of a final physics and dualists who claim that the mental properties are fundamental properties seems to dissolve.[230]

The idea here is that, for all we know, the true physics might take conscious experiences as basic phenomena of our universe. Conscious experiences would then be trivially physical by virtue of being categorized as primitive physical ingredients of our universe.

---

[228] Philips 2013.
[229] Philips 2013.
[230] Montero 1999, p. 192.

This might be taken as a problem in so far as it seems to de-legitimize current investigation into the question of whether materialism is true. For, given our lack of knowledge about the actual physics, we might fail to realize that the actual physics *trivially* implies that materialism is true of our world's conscious experiences, so that in the meantime any arm-chair reasoning to the conclusion that materialism is false is misguided and due solely to being naïve about the actual physics. If we saw the actual physics expressed in the form of an English statement, we'd say, "Ahh, conscious experiences are *primitively* physical, and therefore *trivially* physical; I wish I had known this before I read all those philosophy of mind books and papers." The problem is ultimately that, whereas the mind-body problem is generally taken (at least by those who debate it) to be a substantive issue, the option of the actual physics treating conscious experiences as fundamental is an option according to which actual physics *trivially* implies the conscious experiences—just as trivially as that 4=4 or that every watermelon is a watermelon. Tens of thousands of pages have been written about, say, Jackson's knowledge argument about Mary in the black-and-white room. Jackson initially defended the intuition that, despite knowing all the physics facts, Mary might nevertheless be in no position to deduce what any conscious experiences would be like. Yet, if conscious experiences are primitively physical, she surely could have deduced them. Thus the knowledge argument industry was enabled partly by extreme naivety about what the actual physics is like.

One might suspect that there is also a terminological problem. The idea here is that whether or not the actual physics includes conscious experiences as primitive features of the world is going to be partly terminological. Supposing scientists were to theorize that conscious experiences are primitive features of the world, they would still have to decide whether to call them *physical*. Chomsky expects that they would call them "physical." He says:

We can, however, be fairly sure that there will be a physical explanation for the phenomena in question, if they can be explained at all, for an uninteresting terminological reason, namely that the concept of physical explanation will no doubt be extended to incorporate whatever is discovered in this domain, exactly as it was extended to accommodate gravitational and electromagnetic forces . . . and numerous other entities and processes that would have offended the common sense of earlier generations.[231]

Chomsky's position is plausible. In the past, physicists have expanded their conception of the physical any time they have discovered something new. Thus we might expect this trend to continue even in the case that conscious experiences are taken to be primitive features of the world. Yet there is still the possibility that they call it non-physical. In light of this, it would be presumptuous for us to have arm-chair discussions now about whether materialism is true; for no matter how intelligently we might think about the issue, one's theory might be contradicted in the future by some swift act of dubbing.

Indeed, the terminological problem might seem even more serious: for in so far as a future act of dubbing might be required, it might be that there is currently *no fact of the matter* one way or another whether materialism is true. This would render materialism neither true nor false, which would be even worse than if its truth value was merely difficult to ascertain.

Yet I will argue that the terminological problem can be avoided. All we have to do is to decide, right now, to make certain terminological decisions about what can count as physical. Some philosophers basically do this already, even if they don't think of what they are doing as terminological. For instance, some believe that fundamental mentality should be considered inconsistent with physicalism, so that, if we discover that there is fundamental mentality, it should not count as physical. For instance, Loewer writes:

If it were to turn out that to account for certain clearly physical events physicists needed to posit fundamental intentional, or phenomenal, properties, then the resulting theory would not be physical.[232]

---

[231] Chomsky 1968, p. 83-4.
[232] Loewer 1996, p. 40.

Likewise, Jessica Wilson accepts a "No Fundamental Mentality" constraint on what can count as "physical."[233] And to allow that there can be fundamental mentality is, as Crook and Gillett say, to "abandon . . . the core ideas of physicalism."[234] Thus, if we want, we can decide right now that fundamental mentality is inconsistent with materialism. What we'd effectively have to do, then, is to modify the definition of materialism accordingly. For D_{jac} and D_{cha} would incorporate physics, but not necessarily the entire physics. It would incorporate the entire physics, minus any claims asserting the existence of fundamental mentality. Or else we could decide right now that, if there is fundamental mentality, it *is* going to count as physical. Either way, there is no need to have a terminological problem of the sort whereby the thesis of materialism will lack a truth value until some future date when certain scientists choose to dub one way or another.

Ultimately, then, the only problem having to do with the option that there might be fundamental mentality is as follows. If conscious experiences are fundamental, then this, in conjunction with a decision about whether to call them physical, determines whether materialism is true; so that, any fancy thinking we might do about whether materialism is true is liable to go awry simply by virtue of our naivety about the relevant dubbing.

In sum, *Currentism* fails to capture the philosophical intent of the question of materialism, while *Futurism* is an option according to which the question of materialism cannot be solved from the arm-chair. Yet, although these are serious problems for the traditional materialist debate, I will now argue that these problems are not problems for Chalmers' two-dimensional argument.

Currentism, as always, is a non-starter; so I will ignore it. Therefore materialism should be defined in a *Futurist* way. Yet on the *Futurist* option, it is supposed to be a

---

[233] Wilson 2006.
[234] Crook and Gillett 2001, p. 345.

133

problem that we don't know the true physics. How serious of a problem should this be for Chalmers' argument? There are three potential sources of trouble, but I will argue that none of them are problematic for Chalmers' argument.

First, there is the worry that, in so far as we don't know the true physics, we don't quite understand what we're asking when we ask whether materialism is true. For the thesis of materialism is roughly a thesis about what the true physics entails. If we don't know the true physics, then asking whether the true physics entails such and such would be like asking whether John can hit a home run without knowing who John is, or without knowing enough about John. The result is a serious epistemological problem that prevents an argument one way or another from being plausible. Fortunately, this is not a problem for Chalmers' two-dimensional argument, as the *Conceivability Premise* and CP principles are the only potential weaknesses of that argument, and their justification does not depend on the specifics of the structure and dynamics, or the specifics of the intrinsic natures. So, since there is nothing else to the physics besides structure and dynamics or intrinsic natures, there is no aspect of the true physics whose precise details are apt to make Chalmers' two-dimensional argument any less plausible.

Second, there is the worry that, in so far as we don't know the true physics, we might be radically misconceiving the nature of the physics or the nature of consciousness, so that the concepts that we employ in our arm-chair adventures might not line up in a neat way with the actual nature of our world. In this case, the problem is that (as conceived) there is no fact of the matter about whether everything is physical. Yet here again this is not a problem for Chalmers' argument. For, on the one hand, structure is a category that applies to anything, no matter what the structure is "of" exactly. So, there's no chance that, by conceiving of structure, we will

thereby have applied an inapplicable category to our world. On the other hand, the justification for the *Conceivability Premise* and CP principles don't make any assumptions about what there might be, if anything, besides structure and dynamics (i.e., changing structure). Altogether, then, the justification for Chalmers' two-dimensional argument does not depend even partly on premises that might be affected by this objection. So, there is no problem here for Chalmers' two-dimensional argument.

Third and finally, there is the worry that conscious experiences might be fundamental features of our world, in which case materialism would be true or false of our conscious experiences, depending on whether we commit (right now) to calling them physical or non-physical. Yet this cannot be a problem for Chalmers' argument; for either way would be an option according to which Russellian monism follows from the physics$_2$, thus either way would be an option according to which Chalmers' conclusion follows from the physics$_2$.

In this chapter, I first argued that D_{jac} and D_{cha} are *prima facie* plausible options for how to define materialism, and that each of them implies $M_@$, which Chalmers needs for his two-dimensional argument. Then I discussed three different kinds of problems having to do with the definition of materialism. First, I discussed Hawthorne's blocker arguments against D_{jac} and D_{cha}, but I argued that his objections are naive by virtue of trying to accommodate the hypothetical possibility of non-physical blockers, which I argued are impossible. Then I rebutted his critique of D_{cha} even further by arguing that restricted materialisms don't require a restricted definition of materialism anyway, so that D_{cha} wouldn't have to be suitably modified as $_rD_{cha}$. Second, I discussed Kim's lone ammonium molecule argument against D_{jac} and D_{cha}, but I argued that it is unsound by virtue of trying to pack more into the definition of materialism than most materialist's concepts should want it to capture. For the scope of this dissertation, I think this

discussion of Hawthorne's and Kim's objections is a good start into an investigation of whether

$M_@$ can be derived from a tenable definition of materialism.[235] Finally, I discussed Hempelian

type problems for the traditional materialist debate, but I argued that none of them are problems

for Chalmers' two-dimensional argument.

---

[235] In a more thorough investigation about how to define materialism, we might discuss whether the definition of materialism should be written so as to quantify over the logically possible worlds, or over a more restricted set of metaphysically possible worlds. In the latter case, $M_@$ isn't going to follow from a definition of materialism. Instead, $M_@$ will have to be replaced with something like $M_@'$, the thesis that materialism is true of $w_@$ only if, necessarily, $P_{1',2'}T_{1'} \supset A_{1',2'}$. (Note all the primes, indicating intensions that would range over a more restricted set of metaphysically possible worlds.) But, in chapter 4, we discussed Chalmers' arguments for rejecting *Modal Dualism* (MD), which can be taken as reasons to not define materialism in this way.

CHAPTER 7

CONCLUSION

I have motivated, explicated, evaluated, and defended Chalmers' two-dimensional argument against materialism. I will now review the most important aspects of my defense, the aspects that constitute new contributions to the literature.

First, in chapter 4 I corrected Chalmers' claim that his CP principles require ¬MD. I argued instead that any kind of modal dualism that implies counterexamples to Chalmers' CP principles must imply SN, but that SN doesn't imply MD, and therefore ¬MD doesn't imply ¬SN. Similarly, I also argue in chapter 4, contrary Vaidya (2008), that MD doesn't imply SN. Altogether, then, I argue that there is a lack of relevance between SN and MD.

Second, I argued in chapter 4 that strong necessities clouded by epistemic vagueness cannot imply counterexamples to Chalmers' CP principles. For they would have to be necessarily true when evaluated by 1-intensions, but I argued that this would absurdly require logically arbitrary boundary lines to be logically necessary.

Third, in chapter 4, I diagnosed Timothy Williamson's epistemic view of vagueness as having *The Vague Cookie Cutter Problem*, which, as I pointed out, is equally a problem for Kripke's (1971, 1972/2001) theory of reference.

Fourth, in chapter 4, I argued that ideal negative conceivability is ideal positive conceivability in disguise, and thus Chalmers' two-dimensional argument against materialism should be written in terms of ideal positive conceivability and CP+.

Fifth, I argued in chapter 5 that, contrary to Leuenberger (2008), absences of unfamiliar properties are positively conceivable without indeterminacy, as long as one positively conceives the absence of every sufficiently unfamiliar property all at once.

Sixth, I also argued, contrary to Leuenberger (2008), that ideally positively conceiving with an annotatedly conceived universal "that's all" clause is just as reliable of a guide to possibility as ideally positively conceiving while not annotatedly conceiving anything, as long as the conceiver knows with certainty that the part that is positively conceived is consistent with the part that is annotatedly conceived.

Seventh, I argued in chapter 5 that Frankish's (2007) anti-zombie parity of reasons argument cannot function as a good parity of reasons argument against Chalmers' two-dimensional argument against materialism; for it is invalid, and, if it were to be modified so as to be valid, then it would be plausibly unsound.

Eighth, I argued in chapter 6 that non-physical blockers are impossible, in which case the blocker arguments of Hawthorne (2002) and Leuenberger (2008) are moot.

Ninth, I argued in chapter 6, contrary to Hawthorne (2002), that PD* is a strong enough physical dependence principle to capture materialist physical dependence commitments.

Tenth, I also argued in chapter 6, contrary to Hawthorne (2002), that there is no need to give separate definitions for materialism of a world, and for "restricted materialisms."

Eleventh, I argued in chapter 6, contrary to Kim (1987/1993, 1989/1993, 1990/1993), that, since it is an open question in physics whether physical causes are local, a tenable definition of materialism should not have to presuppose that mind-body supervenience is local.

Twelfth, I argued in chapter 6 that Hempelian problems, although serious for the traditional anti-materialist arguments, are not problems for Chalmers' two-dimensional argument against materialism.

In addition to these twelve advances, I have clarified how Chalmers' two-dimensional argument against materialism goes and why it constitutes an advance over traditional anti-materialist arguments. Moreover, given my provisional defense of Chalmers' argument, his argument seems plausible, as all of the criticisms seem implausible or unsound, and the potential problems that remain seem tenuous.

Nevertheless, Chalmers' two-dimensional argument against materialism does not solve the mind-body problem; it does not even attempt to solve the mind-body problem. It merely defends a constraint on the truth of materialism: namely, that materialism is true only if Russellian monism is true. Moreover, it doesn't demonstrate this; it merely renders it plausible. Yet the plausibly of Chalmers' conclusion should be of interest to philosophers, brain scientists, physicists, and anyone else who is curious about the relationship between consciousness and physics. For it basically says that *physics*-ism is true only if non-relational properties of intrinsic natures play a key role in determining the existence and nature of consciousness. Thus it implies that solving the mind-body problem would require discoveries of a novel sort: namely, discoveries of non-relational properties of intrinsic natures. Thus Chalmers' two-dimensional argument against materialism makes it plausible that, if scientists wish to solve the mind-body problem, they should look for creative ways to discover non-relational properties, as there cannot be an illuminating theory of consciousness that only describes relations.

REFERENCES

Alter, Torin. 2006. On the Conditional Analysis of Phenomenal Concepts. *Philosophical Studies*. Volume 131, Number 3, p. 235-53.

Alter, Torin and Robert Howell. 2009. *A Dialogue on Consciousness*. New York: Oxford University Press.

Armstrong, David. 1991. The Causal Theory of Mind. In *The Nature of Mind*. Edited by David Rosenthal. New York: Oxford University Press.

Ayer, A. J. 1936/1952. *Language, Truth, and Logic*. Second Edition. Dover.

Benacerraf, Paul. 1962. Tasks, Super-Tasks, and Modern Eleatics. *Journal of Philosophy*, Volume LIX, p. 765–784.

Black, Max. 1950-1. Achilles and the Tortoise. *Analysis*. Volume XI, p. 91–101.

--- 1952. The Identity of Indiscernibles. *Mind*. Volume 61, Number 242, p. 153-64.

Blackburn, Simon. 1990. Filling in Space. *Analysis*. Volume 50, Number 2, p. 62-5.

Broad, C. D. 1925. *The Mind and its Place in Nature*. Routledge and Kegan Paul.

Brown, Richard. 2010. Deprioritizing the A Priori Arguments Against Physicalism. *Journal of Consciousness Studies*. Volume 17, Number 3-4, p. 47-69.

--- 2012. The Two-Dimensional Argument Against Non-Materialism. http://onemorebrown.wordpress.com/2012/04/13/the-2d-argument-against-non-materialism/

Brüntrup, Godehard. 2011. Panpsychism and Structural Realism. In *The Mental as Fundamental: New Perspectives on Panpsychism*. Edited by Michael Blamauer. Transaction Books: Piscataway, NJ.

Burke, Michael. 2000. The Staccato Run: a Contemporary Issue in the Zenonian Tradition. *The Modern Schoolman*. Volume 78, Number 1, p. 1–8.

Carruthers, Peter. 2004. *The Nature of the Mind: an introduction*. Routledge.

Chalmers, David. 1995. Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies.* Volume 2, Number 3, p. 200-19. As included in Chalmers (2010).

--- 1996. *The Conscious Mind*. Oxford: Oxford University Press.

--- 2002a. Does Conceivability Entail Possibility? In *Conceivability and Possibility*. Edited by Tamar Szabo Gendler & John Hawthorne. (Oxford University Press, p. 145-200)

--- 2002b/2010. Consciousness and Its Place in Nature. In *Philosophy of Mind: Classical and Contemporary Readings*. Edited by David Chalmers. Oxford.

--- 2002c. On Sense and Intension. *Philosophical Perspectives*. Volume 16.

--- 2005. The Two-Dimensional Argument Against Materialism (online draft). http://consc.net/papers/2dargument.html.

--- 2006. Foundations of Two-Dimensional Semantics In *Two-Dimensional Semantics: Foundations and Applications*. Edited by Manuel Garcia-Carpintero & Josep Macia. Oxford: Oxford University Press.

--- 2009a. The Two-Dimensional Argument Against Materialism (abridged version of his (2005) article. In *The Oxford Handbook of the Philosophy of Mind*. Edited by Brian McLaughlin.

--- 2009b. Appendix: Two-Dimensional Semantics. In *Oxford Handbook of the Philosophy of Language*. Edited by Ernest Lepore & Barry Smith. Oxford University Press.

--- 2010. The Two-Dimensional Argument Against Materialism. In David Chalmers' *The Character of Consciousness*. Oxford University Press, p. 141-91.

--- 2011. The Nature of Epistemic Space. In *Epistemic Modality*. Edited by Andy Egan and Brian Weatherson. Oxford: Oxford University Press.

Chalmers, David and Frank Jackson. 2001. Conceptual Analysis and Reductive Explanation. *Philosophical Review* Volume 110, Number 3, p. 315-61.

Chomsky, Noam. 1968. *Language and Mind*. New York: Harcourt Brace and World.

Crane, Tim and Hugh Mellor. 1990. There is No Question of Physicalism. *Mind*. Volume 99, p. 185-206.

Crook, Seth and Carl Gillett. 2001. Why Physics Alone Can't Define the 'Physical': Materialism, Metaphysics, and the Formulation of Physicalism. *Canadian Journal of Philosophy*. Volume 31, Number 3, September, p. 333-359.

Dennett, Daniel C. 1988. Quining Qualia. In *Consciousness in Modern Science*. Edited by Anthony J. Marcel and E. Bisiach. Oxford University Press.

--- 1991. *Consciousness Explained*. Back Bay Books: Boston.

Descartes. 1641a/2009. *Meditations*. In *Modern Philosophy: An Anthology of Primary Sources*, 2nd edition. Edited by Roger Ariew and Eric Watkins. Indianapolis: Hackett.

--- 1641b/2009. Objections and Replies to Meditations. In *Modern Philosophy: An Anthology of Primary Sources*, 2nd edition. Edited by Roger Ariew and Eric Watkins. Indianapolis: Hackett.

Earman, John, and John D. Norton. 1993. Forever Is a Day: Supertasks in Pitowsky and Malament-Hogarth Spacetimes. *Philosophy of Science*. Volume 60, p. 22–42.

--- 1996. Infinite Pains: The Trouble with Supertasks. In *Benacerraf and His Critics*. Edited by Adam Morton and Stephen P. Stich, Oxford: Blackwell, p. 231–261.

Eddington, Sir Arthur. 1920. *Space, Time, and Gravitation*. Cambridge University Press.

Ellis, Brian. 2007. *Scientific Essentialism*. Cambridge University Press.

Feigl, Herbert. 1967. *The Mental and the Physical: The Essay and a Postscript*. Copp Clark Publishing Company: Toronto.

Frankish, Keith. 2007. The Anti-Zombie Argument. *Philosophical Quarterly*. Volume 57, Number 229, p. 650–666.

Gale, Richard M. 1973. O'Conner on the Identity of Indiscernibles. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*. Volume 24, Number 6, November, p. 412-5.

Hawthorne, John. 2002. Blocking Definitions of Materialism. *Philosophical Studies*. Volume 110, Number 2, p. 1-3-113. DOI: 10.1023/A:102020021394.

Hempel, C. 1969. Reduction: Ontological and Linguistic Facets. In *Essays in Honor of Ernest Nagel***.** Edited Sidney Morgenbesser, *et al*. New York: St Martin's Press.

Hill, Christopher and Brian McLaughlin. 1999. There are fewer things in reality than are dreamt of in Chalmers' philosophy. *Philosophy and Phenomenological Research*. Volume 59, Number 2, June.

Hofstadter, Douglas R. 1999. *Godel, Escher, Bach: an Eternal Golden Braid*. Basic Books: New York.

Howell, Robert. 2008. The Two-Dimensionalist Reductio. *Pacific Philosophical Quarterly*. Volume 89, Number 3, p. 348-358.

Hudson, Hud. 1997. Brute Facts. *Australasian Journal of Philosophy*. Volume 75, Number 1, March.

Hume, David. 1739/2009. *A Treatise on Human Nature*. In *Modern Philosophy: An Anthology of Primary Sources*, 2nd edition. Edited by Roger Ariew and Eric Watkins. Indianapolis: Hackett.

Huxley, Thomas Henry and W. J. Youmans. 1868. *The Elements of Physiology and Hygiene: A Text-book for Educational Institutions*. Appleton & Co.

Jackson, Frank. 1982. Epiphenomenal Qualia. *The Philosophical Quarterly*, Volume 32, Number 127, April, p. 127-136.

    --- 1986. What Mary Didn't Know. *The Journal of Philosophy*. Volume 83, Number 5, May, p. 291-95.

    --- 2000. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford University Press.

Kallestrup, Jesper. 2006. Physicalism, Conceivability, and Strong Necessities. *Synthese*. Volume 151, Number 2, p. 273-295. DOI: 10.1007/s11229-004-7325-9.

Kim, Jaegwon. 1987/1993. 'Strong' and 'Global' Supervenience Revisited. *Philosophy and Phenomenological Research.* Volume 48, p. 315–26. Reprinted in Jaegwon Kim (1993) *Supervenience and Mind*. Cambridge University Press, p. 79-91. All page numbers cited are from the reprinted edition.

    --- 1989/1993. The Myth of Nonreductive Materialism. *Proceedings and Addresses of the American Philosophical Association*. Volume 63, Number 3, November edition, p. 31-47. Reprinted in *Supervenience and Mind*. Cambridge University Press, p. 265-284. All page numbers cited are from the reprinted edition.

    --- 1990/1993. Supervenience as a Philosophical Concept. *Metaphilosophy.* Volume 21, p. 1–27. Reprinted in Jaegwon Kim (1993) *Supervenience and Mind*. Cambridge University Press, p. 131-160. All page numbers cited are from the reprinted edition.

    --- 2005. *Physicalism, or Something Near Enough*. Princeton University Press.

Kosslyn, Stephen. 2003. What Shape Are a German Shepherd's Ears? In *The New Humanists: Science at the Edge*. Edited by John Brockman. New York: Barnes & Noble Books.

Kripke, Saul. 1971. Identity and Necessity. In *Identity and Individuation*. Edited by Milton K. Munitz. New York University Press.

    --- 1972/2001. *Naming and Necessity*. Harvard University Press.

Leibniz, Gottfried Wilhelm. 1714/2009. *Monadology*. In *Modern Philosophy: An Anthology of Primary Sources*, 2nd edition. Edited by Roger Ariew and Eric Watkins. Indianapolis: Hackett.

Leuenberger, Stephan. 2008. *Ceteris Absentibus* Physicalism. In *Oxford Studies in Metaphysics*. Edited by Dean Zimmerman. Volume 4.

Levin, Janet. 2008. Taking type-b materialism seriously. *Mind and Language*. Volume 23, Number 4, p. 402-25.

Levine, Joseph. 1983. Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly*. Volume 64, p. 354-61.

--- 2001. *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press.

Lewis, David. 1986. *On the Plurality of Worlds*. Blackwell.

Locke, John. 1690/2009. *An Essay Concerning Human Understanding*. In *Modern Philosophy: An Anthology of Primary Sources*, 2nd edition. Edited by Roger Ariew and Eric Watkins. Indianapolis: Hackett.

Lockwood, Michael. 1989. *Mind, Brain, and the Quantum*. Oxford: Blackwell.

--- 1999. Interview with Stephen Jones: http://www.culture.com.au/brain_proj/CONTENT/LOCK_FRM.HTM Another link: http://www.scribd.com/archive/plans?doc=30328018

Loewer, Barry. 1996. Humean Supervenience. *Philosophical Topics*. Volume 24, p. 101-123.

Lycan, William. 2012. Is property dualism better off than substance dualism? Philosophical Studies. DOI: 10.1007/s11098-012-9867-x.

--- 2006. The Meaning of "Water": An Unsolved Problem. *Philosophical Issues*. Volume 16, Number 1, p. 184-199.

Malebranche, Nicolas. 1674-5/2009. *The Search After Truth*. In *Modern Philosophy: An Anthology of Primary Sources*, 2nd edition. Edited by Roger Ariew and Eric Watkins. Indianapolis: Hackett.

Maxwell, Grover. 1978. Rigid designators and mind-brain identity. In (C.W. Savage, ed.) *Perception and Cognition: Issues in the Foundations of Psychology*. Minnesota Studies in the Philosophy of Science, Volume 9. Minneapolis: University of Minnesota Press.

McGinn, Colin. 1989. Can We Solve The Mind-Body Problem? *Mind*. Volume 98, July edition, p. 349-66.

--- 1991. *The Problem of Consciousness*. Oxford: Blackwell.

--- 2007. McGinn: *Out of Body, Out of Mind*. In McGinn's Minds & Bodies: Philosophers and Their Ideas. Oxford: Oxford University Press.

Melnyck, Andrew. 2003. *A Physicalist Manifesto*. Cambridge University Press.

Montero, Barbara. 1999. The Body Problem. Nous 33:2. 183-200.

Nagasawa, Yujin. 2002. The Knowledge Argument Against Dualism. *Theoria*. Volume 63, p. 205-23.

Nagel, Thomas. 1974. What Is It Like to Be a Bat? *Philosophical Review* 83 (October): p. 435-50.

--- 1986. *The View From Nowhere*. Oxford University Press.

Papineau, David. 2002. *Thinking About Consciousness*. New York: Oxford University Press.

Paul, L. A. 2006. In Defense of Essentialism. *Philosophical Perspectives*. Volume 20.

Paull, R. Cranston and Theodore Sider. 1992. In Defense of Global Supervenience. *Philosophy and Phenomenological Research* Volume 52, Number 4: 833-54.

Philips, Michael. 2013. What is Materialism? *Philosophy Now*. March/April edition.

Place, U. T. 1956. Is Consciousness a Brain Process? British Journal of Psychology. Volume 47, Issue 1, February, p. 44-50.

Popper, Karl & John Eccles. 1977. *The Self and Its Brain: An Argument for Interactionism*. Springer.

Pruss, Alexander. 2001. The Cardinality Objection to David Lewis's Modal Realism. *Philosophical Studies* (May). Volume 104, Issue 2, p. 169-78.

Putnam, Hilary. 1975/1985. The Meaning of Meaning. *In Philosophical Papers, Vol. 2: Mind, Language and Reality*. Cambridge University Press.

Quine, Willard Van Orman. 1953/1980. Two Dogmas of Empiricism. In *From a Logical Point of View*. Second Revised Edition. Harvard University Press.

--- 1960. *Word and Object*. MIT Press.

Rey, Georges. 1983. A Reason for Doubting the Existence of Consciousness. In *Consciousness and Self-Regulation*. Edited by R. Davidson, R. Schwartz and D. Shapiro. Volume 3. New York, Plenum: p. 1–39.

Robertson, Teresa. 2008. Arguments for Origin Essentialism. Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/entries/essential-accidental/origin-essentialism.html

Rosenberg, Greg. 2004. *A Place for Consciousness. Probing the Deep Structure of the Natural World*. Oxford: Oxford University Press.

Russell, Bertrand. 1920/1993. *Introduction to Mathematical Philosophy 2ⁿᵈ Edition*. Dover: New York.

   --- 1921. *The Analysis of Mind.* Allen & Unwin.

   --- 1924. *The ABC of Atoms*. New York: E.P. Dutton & Co.

   --- 1927a. *The Analysis of Matter*. London: Kegan Paul.

   --- 1927b/2001. *An Outline of Philosophy.* New York: Routledge.

   --- 1948. *Human Knowledge: Its Scope and Limits.* New York: Simon and Schuster.

   --- 1958/1985. *The ABC of Relativity*. London: Allen & Unwin.

   --- 1959/1997. *My Philosophical Development*. London: Allen & Unwin.

Ryle, Gilbert. 1949. *The Concept of Mind*. University of Chicago Press: Chicago.

Salmon, Nathan. 1981. *Reference and Essence*. Princeton: Princeton University Press.

   --- 1989. The Logic of What Might Have Been. *The Philosophical Review*. Volume 98, Number 1, p. 3-34.

Searle, John. 1984. *Minds, Brains and Science*. Cambridge: Harvard University Press.

   ---   1992. *The Rediscovery of the Mind*. London: The MIT Press.

Sellars, Wilfrid. 1962. *Frontiers of Science and Philosophy*, edited by Robert Colodny. Pittsburgh: University of Pittsburgh Press, 19: 35-78. Reprinted in *Science, Perception and Reality* (1963). Also available online: http://www.ditext.com/sellars/psim.html#5

Smart, J. J. C. 1959. Sensations and Brain Processes. *The Philosophical Review*. Volume 68, Number 2, April, p. 141-56.

Soames, Scott. 2007. *Reference and Description: The Case against Two-Dimensionalism*. Princeton University Press.

Sprague, Elmer. 1966. The Mind-Brain Problem. In *Dimensions of Mind*. Edited by Sidney Hook. New York: Collier Books.

Sprigge, Timothy L. S. and Alan Montefiore. 1971. Final Causes. *Proceedings of the Aristotelian Society, Supplementary Volumes*. Volume 45, p. 149-92.

Stalnaker, Robert. 1996. Varieties of Supervenience. *Nous*, Volume 30, Supplement: *Philosophical Perspectives*, 10, Metaphysics, p. 221-241.

Stoljar, Daniel. 2001. Two Conceptions of the Physical. *Philosophy and Phenomenological Research.* Volume 62, p. 253-81.

   --- 2009. Physicalism. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/physicalism/

Tegmark, Max. 2008. The Mathematical Universe. *Foundations of Physics*. Volume 38, p. 101-150.

   --- 2008. *Discover* magazine interview: "Is the Universe Actually Made of Math?" http://discovermagazine.com/2008/jul/16-is-the-universe-actually-made-of-math

Thomson, James F. 1954. Tasks and Super-Tasks. *Analysis.* Volume 15, Number 1, p. 1–13. DOI: 10.2307/3326643.

Tooley, Michael. 1972. Abortion and Infanticide. *Philosophy and Public Affairs*. Volume 2.1, p. 37-65. Blackwell Publishers.

Unger, Peter. 1998. The Mystery of the Physical and the Matter of Qualities: A Paper for Professor Shaffer. *Midwest Studies in Philosophy*. Volume 22.

Vaidya, Anand. 2008. Modal Rationalism and Modal Monism. *Erkenntnis*. Volume 68, number 2, p. 191-212. DOI: 10.1007/s10670-007-9093-7.

van Fraassen, Bas C. 1980. *The Scientific Image*. Reprint edition. Oxford University Press.

   --- 2006. Structure: Its Shadow and Substance. *British Journal for the Philosophy of Science*. Volume 57, Number 2, p. 275-307.

van Inwagen, Peter. 1983. *An Essay on Free Will.* Oxford: Oxford University Press.

Whitehead, Alfred North. 1929/1978. *Process and Reality: An Essay in Cosmology*. Revised edition by David Ray Griffin and Donald W. Sherburne. Free Press: New York.

Williamson, Timothy. 1994. *Vagueness*. Routledge: London.

Wills, Adrian. 1999. Herophilus, Erasistratus, and the birth of neuroscience. *The Lancet*. Volume 354, November 13, p. 1719-20.

Wilson, Jessica. 2006. On Characterizing the Physical. *Philosophical Studies*. Volume 131, Number 1, p. 61-99.

Wittgenstein, Ludwig. 1921. *Logisch-Philosophische Abhandlung*. English translations started coming out in 1922: they all call it the *Tractatus Logico-Philosophicus*.

Yablo, Stephen. 1999. Concepts and consciousness. *Philosophy and Phenomenological Research*. Volume 59, Number 2, p. 455-63.