EPISTEMIC GAPS AND THE MIND-BODY PROBLEM

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Thomas Scott Foerster

May 2019

EPISTEMIC GAPS AND THE MIND-BODY PROBLEM

Thomas Scott Foerster, Ph.D.

Cornell University 2019

This dissertation defends materialism from the epistemic arguments against materialism. Materialism is the view that everything is ultimately physical. The epistemic arguments against materialism claim that there is an epistemic gap between physical and phenomenal truths (for example, that knowing the physical truths does not put you in a position to know the phenomenal truths), and conclude from this that there is a corresponding gap in the world between physical and phenomenal truths, and materialism is false.

Chapter 1 introduces materialism and the arguments against materialism that I respond to in this dissertation.

Chapters 2 and 3 explore the phenomenal concept strategy (PCS), one promising materialist response to the epistemic arguments against materialism. The PCS admits that there is an epistemic gap between physical and phenomenal truths, but claims that it arises because of something special about our *concepts* of consciousness and not because of something special about consciousness itself.

Chapter 4 considers what reason there is to think that there is an epistemic gap between physical and phenomenal truths in the first place. I argue that some arguments might succeed in showing that there is such an epistemic gap, but that if these arguments succeed then they end up vindicating the PCS.

Chapter 5 considers more generally what is required for an epistemic gap to exist. In particular, it considers whether conceptual analysis is necessary to close an epistemic gap.

Biographical Sketch

Thomas Scott Foerster was born in London, Ontario, Canada. He received the Bachelor of Arts in Philosophy from the University of Toronto in 2012 and the Master of Arts in Philosophy from Cornell University in 2016.

Acknowledgments

For helpful discussion and comments, thanks to my committee members: Nico Silins, Will Starr, and especially my chair Derk Pereboom. Thanks also to participants of a 2016 Cornell Philosophy workshop for comments on an ancestor of chapter 2. And thanks to two anonymous referees for comments on a version of chapter 4.

Contents

Chapter 1

Introduction

1.1 Consciousness

Like my cat, I am conscious: there is something it is like to be me. So are you. We can experience pleasure, feel nauseated, see green, hear barking, and experience déjà vu. It is not just that we can behave in certain ways, or process information, or instantiate various functional states, though we can do all these things. It is that it *feels* a certain way to do them. If we were not conscious, then it would be all dark inside. Consciousness is a central and important feature of our lives.[1]

Many philosophers find consciousness mysterious. On the one hand, we seem to have good reason to think that everything, including consciousness, is ultimately physical. On the other hand, it is hard to see how consciousness could be just physical. For example, one natural thought is that consciousness

[1]Terminological note: my use of the term 'consciousness' differs from how it is used in ordinary speech. Following standard philosophical use, I say that a person is conscious iff there is something it is like to be that person, and a mental state is conscious iff there is something it is like to be in that mental state. So consciousness, in this sense, is not the same as wakefulness or awareness. For example, you count as conscious if you are asleep and dreaming, since there is something it is like to dream.

is just brain activity of some sort. But why isn't all that brain activity going on without any consciousness? There does not seem to be any necessary connection between the two. As McGinn (1989, p. 349) famously put it, "How can technicolor phenomenology arise from soggy grey matter?"

So we get pulled in two different directions. Philosophers of mind accordingly divide into two main camps. Roughly, *materialists* about the mind think that consciousness is physical, while *dualists* think that consciousness is at least partly non-physical. This dissertation is about the arguments that dualists make against materialism, and how materialists should respond to them. In the rest of this chapter, I will explain materialism and dualism more precisely, review the leading epistemic arguments against materialism, and then spell out in more detail how I think materialists should respond to them.

To be clear, although I accept materialism, I will not positively argue for it in this dissertation. My aim is just to show that the leading arguments *against* materialism are unsuccessful. That will be more than enough work.

1.2 Materialism and dualism

Materialism (or "physicalism") is the view that all truths are either physical or grounded in the physical truths.[2] The idea, as it is sometimes put, is that once God settled all the physical truths, he did not have any more work to do to make the world the way it is. This requires, at the very least, that all truths are necessitated by the physical truths. Materialism about the mind is the view that

[2]To say that A is grounded in B is to say, roughly, that A holds in virtue of B, or that A is nothing over and above B. For example, the fragility of a vase is grounded in its molecular make-up, and the fact that my hat is colored is grounded in the fact that it is red. See Rosen (2010), Audi (2012) and Bliss and Trogdon (2014) for introductions to grounding. If you are still uncomfortable with grounding, you can substitute necessitation-talk for grounding-talk in what follows without too much loss.

all *mental* truths—truths, say, about what you are thinking, feeling, or intending—are grounded in the physical truths. Going forward, when I say 'materialism' I will mean 'materialism about the mind'.

Here the physical truths are understood narrowly as truths that can be stated in the language of fundamental physics. To be precise, we can call these '*narrowly* physical truths', though usually I will omit the 'narrowly' qualification. For example, 'there is a hydrogen atom with such-and-such a velocity' counts as a narrowly physical truth, but 'there is a chair in my apartment' does not (though it is presumably *grounded* in the narrowly physical truths). We can call truths that are either narrowly physical or grounded in the narrowly physical truths '*broadly* physical truths'.[3, 4]

Anti-materialism is the negation of materialism. The most popular form of anti-materialism is *dualism*, which says that in addition to the physical truths, there are mental truths that are not grounded in the physical truths. A dualist might claim that I am conscious not simply because of what is going on in my brain, but also because of something going on in a non-physical realm. A different form of anti-materialism is *idealism*, which says that all truths are either mental or grounded in the mental truths.

Materialism has a lot going for it. It is parsimonious, it seems to fit well with the picture of the world given to us by the sciences, and it avoids hard questions about how the mental and physical interact if the mental is neither physical nor grounded in the physical. See Stoljar (2015, §17) and Papineau (2002, chapter 1)

[3]The distinction between broadly and narrowly physical truths is from Chalmers (2010, p. 192). In a later paper, Chalmers (2015) uses the same labels to draw a different distinction. I will always use 'narrowly physical' and 'broadly physical' as defined above.

[4]I am setting aside some complications in characterizing materialism. For example, one complication is whether the the relevant physical truths those of *actual* physics or those of a *completed* physics. See Stoljar (2010) for an in-depth discussion of this and other complications, and Goff (2017, §2.1) for a more recent overview. These complications are not important for anything I am going to go on to say in this dissertation.

for an overview of the case for materialism. But as I noted above, I will not try work out a rigorous argument for materialism in this dissertation. Instead, my focus is on arguments *against* materialism.

1.3 Arguments against materialism

Two arguments against materialism have received the most attention: the knowledge argument and the conceivability argument. I will consider each in turn.[5] My aim in this section is just to get the arguments on the table. The assessment of them comes later.

1.3.1 The knowledge argument

The knowledge argument starts by claiming that you can know all the (narrowly)[6] physical truths without being in a position to know all the truths about consciousness (or "phenomenal truths"). For example, Jackson (1982, 1986) has us imagine someone, Mary, who knows all the physical truths but who has never seen red. He claims that, despite knowing all the physical truths, Mary is not in a position to know what it is like to see red. Proponents of the knowledge argument then infer from this that at least some phenomenal truths are not grounded in the physical truths, and materialism is false.

Note that Jackson cannot start by supposing that Mary knows all the *broadly* physical truths, since materialists think that phenomenal truths—along with all other truths—*are* broadly physical. So they think that if Mary knows all the broadly physical truths, then she is trivially in a position to know what it is like

[5] My presentation of these arguments largely follows Chalmers (2003a) and Stoljar (2006).
[6] I usually omit this qualification in what follows, but sometimes include it for clarity.

to see red (provided that there is such a truth for her to know). Of course, Jackson thinks that materialists are wrong about this,[7] but he cannot simply assume at the outset that they are wrong. Supposing that Mary knows all the *narrowly* physical truths avoids this problem, since even materialists think that phenomenal truths are not narrowly physical.[8]

The knowledge argument can be formalized as follows:

(K1) Knowing all the physical truths does not put you in a position to know all the phenomenal truths.

(K2) If (K1) is true, then there are phenomenal truths that are not grounded in the physical truths, and materialism is false.

(K3) ∴ Materialism is false.

It is important to note that (K3) does not follow from (K1) alone. Materialism is a view about what the world is like. It is not a view about knowledge. What we can know is one thing. What the world is like is another thing. So it is open to materialists to admit (K1), but deny that it entails anything interesting about what the world is like. Against this, anti-materialists must claim that facts knowledge sometimes do entail that the world is a particular way. This is what (K2) claims: that the *epistemic* premise (K1) entails the *metaphysical* conclusion (K3).

(K1) is sometimes defended by appeal to intuitions about imagined cases, such as the intuition that Mary would not be able to learn what it is like to see red while in the black-and-white room. A more satisfying defense of (K1) appeals to considerations about physical concepts and concepts of consciousness (or "phenomenal concepts"). For example, some anti-materialists claim that knowing

[7] At least, Jackson *did* think this. He has since become a materialist. See Jackson (2003).

[8] The one possible exception is certain Russellian monists, who I discuss in §1.4.3 below.

all the physical truths would put you in a position to know phenomenal truths only if there were an analysis of phenomenal concepts into physical concepts, but that there is no such analysis. If this is right, then (K1) is true.

Anti-materialists rarely defend (K2) directly. More often, they defend (K2) indirectly by objecting to materialists arguments *against* (K2). For example, anti-materialists sometimes argue that the "old fact/new guise" objection to (K2) fails, and consequently that (K2) is true—since, they assume, this is the only serious objection to (K2).[9]

What about direct defenses of (K2)? Here the natural strategy is to appeal to general principles connecting knowledge and grounding. For example, perhaps anti-materialists can claim that, in general, if knowing all the truths about some domain D_1 (e.g., the physical) does not you in a position to know all the truths about some other domain D_2 (e.g., the phenomenal), then the truths in D_1 do not ground the truths in D_2. This straightforwardly entails (K2). If this is right, then there is a close connection between the structure of our knowledge and the structure of the world.[10]

1.3.2 The conceivability argument

The conceivability argument starts by claiming that it is conceivable that all the physical truths hold without all the phenomenal truths holding. For example, Chalmers (1996, 2010) claims that it is conceivable that there is a *zombie world*: a world physically identical to the actual world in which no one is conscious. Proponents of the conceivability argument then infer from this that it is possible for there to be such a world. If this is right, then the physical truths do not

[9]See Nida-Rümelin (2009, §4.2) for an overview of the old fact/new guise objection.

[10]See Chalmers (2012) for an extended defense of this sort of view. More specifically, he argues that knowing all the fundamental truths (together with some indexical truths) puts you in a position to know every truth.

necessitate all the phenomenal truths, and so some phenomenal truths are not grounded in the physical truths (since grounding requires necessitation). As before, it follows that materialism is false.

The conceivability argument can be formalized as follows:

(C1) It is conceivable that all the physical truths hold without all the phenomenal truths holding.

(C2) If (C1) is true, then there are phenomenal truths that are not grounded in the physical truths, and materialism is false.

(C3) $\therefore$ Materialism is false.

Again, it is important to note that (C3) does not follow from (C1) alone. Materialism is a view about what the world is like. It is not a view about conceivability. What we can conceive is one thing. What the world is like is another thing. So it is open to materialists to admit (C1), but deny that it entails anything interesting about what the world is like. Against this, anti-materialists must claim that facts conceivability sometimes do entail that the world is a particular way. This is what (C2) claims: that the *epistemic* premise (C1) entails the *metaphysical* conclusion (C3).

Like (K1), (C1) is sometimes defended by appeal to intuitions, such as the intuition that a zombie world is conceivable. A more satisfying defense of (C1) appeals to considerations about physical and phenomenal concepts. For example, some anti-materialists claim that a zombie world would be inconceivable only if there were something conceptually incoherent about it, but that there can be no such conceptual incoherence because phenomenal concepts cannot be analyzed into physical concepts.[11]

[11]Chalmers (1996, p. 104) argues for (C1) in this way. I say more about this argument later, in chapter 4.

(C2) is generally defended by appeal to general principles that link conceivability to possibility. For example, one principle claims that if anything is conceivable, then it is also possible. If this is right, then (C2) entails that some phenomenal truths are not necessitated by the physical truths, and so are not grounded in them. A complication arises from the fact that, as stated, there are clear counterexamples to this principle linking conceivability to possibility. (For example, 'water is not H_2O' is conceivable but impossible.) Proponents of the conceivability argument attempt to formulate more nuanced principles that avoid these counterexamples.[12]

1.3.3 The form of the arguments

The knowledge and conceivability arguments differ in some important ways, but they have the same basic form. So in thinking about how materialists can respond to these arguments, it is helpful to abstract from the differences between them and think about their common form. As Chalmers (2003a) notes, both arguments claim that there is a certain *epistemic gap* between the physical and phenomenal truths, and then infer from this that there is a corresponding *ontological gap* between them—a gap in the world between the physical and phenomenal. I will refer to anti-materialist arguments of this form as 'the epistemic arguments'.

Here is how Chalmers formalizes the epistemic arguments:

(E) There is an epistemic gap between physical and phenomenal truths.

(E→O) If there is an epistemic gap between physical and phenomenal truths, then there is an ontological gap, and materialism is false.

[12]Most notably, Chalmers (2010) formulates a conceivability-possibility principle using a framework of two-dimensional semantics.

(O) $\therefore$ Materialism is false. (Chalmers 2003a, p. 208)[13]

The epistemic gap that the knowledge argument talks about is a *knowledge* gap: Mary's being unable to learn what it is like to see red from her knowledge of the physical truths. The epistemic gap that the conceivability argument talks about is a *conceivability* gap: its being conceivable that the physical truths hold without all the phenomenal truths holding. The ontological gap that both arguments talk about is a lack of grounding: the phenomenal truths' being ungrounded in the physical truths.

1.4 Materialist responses

Materialists have a few options in responding to the epistemic arguments. The most straightforward options are to deny (E) or deny (E→O). Type-A materialists take the first option, and type-B materialists take the second option. A more nuanced option, taken by Russellian monists, is to claim that the argument equivocates: on one reading (E) is true, on another reading (E→O) is true, but on no reading that targets an antecedently plausible version of materialism are they *both* true. Finally, it is possible to deny the *conjunction* of (E) and (E→O) without taking a stand on which one is false. This is the response I am most sympathetic to. I explain each response in more detail below.

1.4.1 Type-A materialism

Type-A materialists deny (E). That is, they deny that there is an epistemic gap between physical and phenomenal truths. So they claim that Mary, while in the

[13] I have changed Chalmers' labeling of the premises to make it more mnemonic. ('E' for 'epistemic gap', 'O' for 'ontological gap'.)

black-and-white room, would be able to learn what it is like to see red (provided that there really is such a truth to know), and that zombies are inconceivable. Traditionally, type-A materialists have been either analytic functionalists or eliminativists.

Analytic functionalists claim that phenomenal concepts can be analyzed into physical or functional concepts.[14] To take a toy example, they might claim that it is a conceptual truth that pain is whatever state that tends to be caused by bodily damage and which in turn causes distress. (The actual analysis of the concept <pain> would presumably be much more complicated than this.[15]) If analytic functionalism is true, then so long as there is no epistemic gap between the physical truths and the relevant functional truths, there will also be no epistemic gap between the physical and phenomenal truths.

Eliminativists claim that no one is phenomenally conscious, and so there are no (positive) phenomenal truths. If there are no phenomenal truths, then a fortiori there is no epistemic gap between physical and phenomenal truths. For example, Mary is trivially in a position to know all the phenomenal truths while in the black-and-white room, because there are none for her to know. Eliminativism is a radical view, and few philosophers explicitly endorse it.[16] More often, eliminativism is a view that philosophers are accused of holding by their opponents. For example, analytic functionalists are sometimes accused of being eliminativists.

[14]See, for example, Lewis (1966, 1972) and the references in Levin (2018).

[15]I am glossing over two more complications. First, the analysis in terms of 'bodily damage' is not in (narrowly) physical terms as I understand them above—terms of fundamental physics. Second, there is the mention of distress in the above toy analysis. So the analysis is also partly in terms of mental states. That is allowed by analytic functionalists. They claim, however, that after we get all the analyses of mental states on the table, we can "Ramsify" out the mention of mental states, and so in the end the analysis will not take any mental state concepts as primitive. See Lewis (1994. p. 56), among others, for discussion.

[16]The one exception is Frankish (2016) and other illusionists, who accept that it *seems* like we are conscious, but argue that this is an illusion.

Type-A materialism is currently relatively unpopular.[17] This is because it is widely thought that type-A materialists must be either analytic functionalists or eliminativists,[18] and analytic functionalism and eliminativism themselves are currently relatively unpopular. But although nearly all type-A materialists have been either analytic functionalists or eliminativists, there is a more plausible version of type-A materialism that is not committed to either analytic functionalism or eliminativism. I discuss this version of type-A materialism in chapter 4.

1.4.2 Type-B materialism

Type-B materialists deny (E→O). That is, they deny that the epistemic gap between physical and phenomenal truths entails an ontological gap. So they think that even if Mary is not in a position to know what it is like to see red while in the black-and-white room, this does not entail that there are phenomenal truths that are not grounded in the phenomenal truths. They also think that even if zombies are conceivable, this does not entail that they are possible.

Type-B materialists break into two camps.[19] In the first camp are those who think that epistemic gaps *in general* do not entail ontological gaps. In defense of this claim, they generally argue that there are many epistemic gaps between physical truths and other truths without any corresponding ontological gaps. For example, Block and Stalnaker (1999) argue that knowing the physical truths does

[17] On the 2009 PhilPapers survey, only 16% of respondents accepted or leaned toward the type-A materialist view that zombies are inconceivable, compared to 35% who accepted or leaned toward the type-B materialist view (explained below) that they are conceivable but impossible (Bourget and Chalmers 2014).

[18] Chalmers' (2003a) thorough taxonomy of views in philosophy of mind includes no clear examples of type-A materialists who are neither analytic functionalist or eliminativists, and expresses doubt that there is room for such a view.

[19] Levine (2001, p. 50) calls those in the first camp 'type-NE' (no exception) materialists, and those in the second camp 'type-E' (exception) materialists. (Beware that what Levine calls 'type-E' materialism differs from what Chalmers 2003a calls 'type-E' *dualism* in his taxonomy of views about consciousness.)

not put you in a position to know even ordinary macroscopic truths, such as that the pot is boiling.[20] (Remember that I am using 'physical' to mean 'narrowly physical'. That the pot is boiling does not count as a physical truth in this sense of 'physical'.)

In the second camp are those who grant that that epistemic gaps usually do entail ontological gaps, but who argue that the epistemic gap between physical and phenomenal truths is special in some way and does not entail an ontological gap. In defense of this claim, they generally appeal to some version of the phenomenal concept strategy (PCS). Proponents of the PCS argue that the epistemic gap between physical and phenomenal truths arises not because of something special about consciousness itself, but rather because of something special about our *concepts* of consciousness. This, in turn, is supposed to undermine (E→O).

The PCS has many proponents,[21] but it has also been strongly criticized by Chalmers and others. Much of this dissertation considers challenges to the PCS and assesses how well it can respond to them.

1.4.3 Russellian monism

Russellian monists do not straightforwardly deny either (E) or (E→O). Instead, they claim that 'physical' is ambiguous. On one reading of 'physical', both (E) and (E→O) are true and the argument succeeds, but it refutes only one version of materialism—a version that Russellian monists reject for reasons independent of the epistemic arguments. On the other reading of 'physical', (E) is false and the argument is unsound. Different Russellian monists distinguish between the

[20] Other proponents of this sort of view include Byrne (1999), Levine (1998, 2001), and Schaffer (2017).

[21] These include Loar (1990/7), Tye (2000), Perry (2001), Papineau (2002, 2007), Levin (2007), and Balog (2012b), among others.

two senses of 'physical' in different ways.[22] Here I will work with one simple way of drawing the distinction, which is not necessarily the best.[23]

On the first reading of 'physical', physical properties are understood narrowly to include only dispositions. Call these 'physical−' properties. The causal role that an electron plays then counts as physical−, but the categorical basis of that causal role (if there is any) does not count as physical−. On the second reading of 'physical', physical properties are understood more broadly to include both dispositions and their categorical bases, if they have any. Call these 'physical+' properties. If there are no categorical bases of physical− properties, then every physical+ property is also a physical− property.

Materialism can then be understood in two ways, depending on how 'physical' is understood. One version, materialism−, says that all truths are physical− or grounded in the physical− truths. The other version, materialism+, says that all truths are physical+ or grounded in the physical+ truths. If physical properties have categorical bases, then materialism− is false: truths about the categorical bases of physical properties are grounded in physical+ but not physical− truths. But Russellian monists reject materialism−, because they think that dispositional properties must have categorical bases.

Turning to the epistemic arguments: Russellian monists agree with anti-materialists that there is an epistemic gap between the physical− and phenomenal truths. For example, they think that Mary, if she knew only the physical− truths, would not be in a position to know what it is like to see red. They also accept that epistemic gaps, in general, entail ontological gaps. So they think that on one reading, the epistemic arguments succeed. But on this reading, they refute only materialism−, not materialism+. At the same time, Russellian monists deny

[22]Note that this distinction is completely independent of the distinction between broadly and narrowly physical truths from above.

[23]See Stoljar (2014) for a critical overview of four ways of drawing the distinction.

that there is an epistemic gap between the *physical+* and phenomenal truths. For example, they think that if Mary knew all the physical+ truths, then she *would* be in a position to know what it is like to see red. If this is right, then the the epistemic arguments cannot refute materialism+.

There has been a flurry of interest in Russellian monism recently.[24] But I will have little to say about Russellian monism in this dissertation. My focus is instead on the less metaphysically speculative materialist views: type-A and type-B materialism. This is not because I think that Russellian monism is obviously false (though it faces serious challenges[25]). It is just that Russellian monism is usually a view that materialists retreat to after being persuaded by anti-materialists that type-A and type-B materialism are untenable. And I think that this retreat is unnecessary. As I will argue, type-A and type-B materialism have much more to be said for them than most Russellian monists believe.

1.4.4 Agnosticism

The final response is to deny the conjunction of (E) and (E→O) without taking a stand on which one is false. This amounts to claiming that either type-A or type-B materialism is true, without taking a stand on which one is true. This view, to my knowledge, has not previously been defended by anyone. It might seem like a cop-out—more a statement of faith that the epistemic arguments fail some way or other than a serious response to them. But I think that this is a mistake. There is, I will argue, a major payoff to avoiding tunnel vision and keeping both

[24]See, for example, Pereboom (2011), Alter and Nagasawa (2015), Bruntrup and Jaskolla (2017), and Goff (2017). See also Stoljar (2001) for a classic early defense of Russellian monism. (Note that in a more recent 2015 paper, Stoljar endorses what he calls 'Nagelian monism' over Russellian monism.)

[25]Most notably, there is the so-called "combination problem": the problem of explaining how the categorical bases of physical dispositions combine to form experiences. See Chalmers (2016) for an overview of the combination problem.

type-A and type-B materialism in mind when defending materialism.

Roughly, the payoff is that type-A and type-B materialism each end up picking up the slack of the other: individually each has weaknesses, but together they are strong. A bit more fully, there is a serious objection to type-A materialism—but if it succeeds, then it ends up vindicating a version of type-B materialism. And there is a serious objection to type-B materialism—but if it succeeds, then it ends up vindicating a version of type-A materialism. Materialists could try to challenge each objection on its own terms. But they can avoid taking a stand on a lot of thorny philosophical issues by simply being agnostic about whether the relevant objection succeeds, and being prepared to retreat to the other version of materialism if the objection does succeed.

Of course, the disjunction of type-A and type-B materialism is consistent with both type-A and type-B materialism. It is not a competitor to them. But it is worth having on the table as a separate response to the epistemic arguments, since a defense of the *disjunction* of type-A and type-B materialism might differ from a defense of type-A materialism or a defense of type-B materialism. In fact, as I suggested in the previous paragraph, a defense of the disjunction can side-step philosophical issues that a defense of type-A or a defense of type-B materialism would have to address. This is the line of argument I develop in chapter 4.

1.5 The structure of this dissertation

Much of this dissertation is concerned with the PCS. But the underlying theme is agnosticism between type-A and type-B materialism, and how type-A and type-B materialism can each pick up the slack of the other.

I start, in **chapter 2**, by considering the PCS—again, a popular defense of type-B materialism. Despite the popularity of the PCS, it is surprisingly hard to find a clear statement of what premises it is committed to or how, exactly, it is supposed to undermine the epistemic arguments. So the first order of business is to clarify the PCS. After that, I argue that the leading versions of the PCS are committed to what I call 'the Analysis Principle': roughly, the claim that conceptual analysis is necessary to close epistemic gaps. I then conditionally defend these versions of the PCS, arguing that *if* the Analysis Principle (which I take no stand on in chapter 2) is true, then they succeed.

Chapter 3 responds to a dilemma posed by Chalmers which aims to show that no version of the PCS can succeed. I argue that he is wrong.

What if the Analysis Principle, which the leading versions of the PCS rely on, is false? Then, **chapter 4** argues, the epistemic arguments fail for reasons independent of the PCS. This is because anti-materialists themselves must appeal to the Analysis Principle in defending (E), the claim that there is an epistemic gap. Specifically, I argue that anti-materialists can defend (E) only by making the following argument: (i) phenomenal concepts cannot be analyzed into physical concepts, and so by the Analysis Principle (ii) there is an epistemic gap between physical and phenomenal truths. If the Analysis Principle is false, then this argument is invalid, and the door is open to type-A materialism.

Chapter 5 assesses the Analysis Principle. I try to develop the best argument for it, but ultimately side with Chalmers and Jackson (2001) in rejecting it.

Chapter 2

The phenomenal concept strategy

2.1 Introduction

As we saw in chapter 1, the epistemic arguments have two premises. The first premise, (E), claims that there is an epistemic gap between the physical and phenomenal truths (e.g., that zombies are conceivable). The second premise, (E→O), claims that this epistemic gap entails a corresponding ontological gap (e.g., that the conceivability of zombies entails their possibility). In response, materialists generally either deny (E) or deny (E→O). That is, they either deny that there is an epistemic gap between the physical and phenomenal truths, or deny that this epistemic gap entails a corresponding ontological gap.

Proponents of the phenomenal concept strategy (PCS) take the second option: they admit that there is an epistemic gap, but deny that it entails an ontological gap. They argue that the epistemic gap arises not because of something special about consciousness itself, but rather because of something special about our *concepts* of consciousness (or "phenomenal concepts"). They add that these special features of our concepts of consciousness are no evidence of an ontological

gap. From these claims, they conclude the epistemic gap is no evidence of an ontological gap. This does not show that (E→O) is false, but it significantly reduces its plausibility.[1]

The PCS is a popular response to anti-materialist arguments. Versions of the PCS have been defended by Loar (1990/7), Levine (1998), Tye (2000), Perry (2001), Papineau (2002, 2007), Levin (2007), and Balog (2012b), among others. But it is surprisingly hard to find a clear statement of what premises, exactly, the PCS is committed to or how its conclusion is supposed to undermine the epistemic arguments. On top of this, proponents of the PCS often take it to be committed to unnecessarily strong premises, or an unnecessarily weak conclusion. As a result, the full force of the PCS has been under-appreciated—even by its proponents.[2]

In this chapter I will clarify the PCS and highlight what I take to be the most promising version of it. I will start by giving a general overview of the PCS and explaining how, exactly, its conclusion undermines the epistemic arguments. Then I will assess various versions of the PCS, conditionally defending what I call 'the analyzability version'. My defense is conditional in that I argue merely that *if* a certain principle is true, then the analyzability version of the PCS succeeds. But as I explain below, even this modest conclusion is enough to undermine the epistemic arguments, since anti-materialists themselves must rely on the relevant principle to argue against materialism.

[1] At least, this is how I understand the PCS. Other accounts of it differ somewhat from mine. On most other accounts, its conclusion is that the epistemic gap can be physically explained. See §2.3 for discussion.

[2] One prominent defender of the PCS claimed recently that there is no non-question-begging way for materialists to defend the PCS (Balogb 2012, pp. 19-20). This, I believe, seriously underestimates the PCS.

2.2 Overview of the PCS

On my account, the PCS consists in three main claims. First, a claim about phenomenal concepts: that they have a certain property, ϕ. Second, a claim about explanation: that the fact that phenomenal concepts are ϕ explains the epistemic gap. Third, a claim about evidence: that the fact that phenomenal concepts are ϕ is no evidence of an ontological gap. Proponents of the PCS then infer from these claims that the epistemic gap is no evidence of an ontological gap. In the next section I consider how this conclusion is supposed to undermine the epistemic arguments. (Although proponents of the PCS deny (E$\rightarrow$O), they do not argue against it directly.)

So, for some ϕ, proponents of the PCS claim:

(1) *Attribution:* Phenomenal concepts are ϕ.

(2) *Explanation:* The fact that phenomenal concepts are ϕ explains the epistemic gap.

(3) *No evidence:* The fact that phenomenal concepts are ϕ is no evidence of an ontological gap.

From (1)-(3), they conclude:

(4) The epistemic gap is no evidence of an ontological gap.

By themselves, (1)-(3) do not logically entail (4). For now we can add a bridge premise to make the argument trivially valid:

Bridge: (1)-(3) are true, then (4) is true.

Below I will consider whether the bridge premise can be justified by appeal to general principles governing evidence and explanation.

This is just an argument schema. Different versions of the PCS substitute different properties for ϕ—say, the property of being recognitional, or quotational, or non-analyzable into physical or functional concepts. Below I will consider these and other versions of the PCS, conditionally defending the version that claims that phenomenal concepts are non-analyzable into physical or functional concepts. But I will start by clarifying a few general points about the PCS and my account of it.

First, (2) and (4) talk ambiguously and somewhat misleadingly about "the" epistemic gap. The knowledge and conceivability arguments talk about two *different* epistemic gaps. In principle, these epistemic gaps could be explained in two different ways. So there could be one version of the PCS that targets the knowledge gap and a different version that targets the conceivability gap. But in practice, most versions of the PCS aim to show that none of the relevant epistemic gaps is evidence of an ontological gap. In any case, since the two epistemic gaps are so closely related, it is reasonable to expect that whatever explains one will also explain the other.

Second, note the importance of (3), which proponents of the PCS often fail to make explicit. By themselves, (1) and (2) do nothing to help materialists. Even anti-materialists can admit that phenomenal concepts have some property that explains the epistemic gap—say, the property of referring to non-physical states. That would explain the epistemic gap, but it would also entail that there is an ontological gap. So what matters for proponents of the PCS is not merely whether the epistemic gap can be explained, but also *how* it can be explained.

Third, on my account the PCS is not explicitly committed to what Levine (2007) calls 'the materialist constraint': the claim that there is a physical explanation of the fact that phenomenal concepts are ϕ.[3] It might be argued that (3) is

[3]Levine (2007), Levin (2007), and Chalmers (2007) all take the PCS to be committed to the

true only if there is a physical explanation of the fact that phenomenal concepts are ϕ. If this is right, then even on my account the PCS is *implicitly* committed to the materialist constraint. Still, it is open to proponents of the PCS to at least try to defend (3) without giving a physical explanation of the fact that phenomenal concepts are ϕ.[4] So it is best not to build the materialist constraint into an account of the PCS.

Fourth, although (3) claims that the epistemic gap is *no* evidence of an ontological gap, the PCS would still work if this claim were weakened somewhat. For example, (3) could be modified to say that the fact that phenomenal concepts are ϕ is at most weak evidence of an ontological gap. Then the conclusion would have to be modified to say that the epistemic gap is at most weak evidence of an ontological gap. But even this conclusion significantly undermines the epistemic arguments. I believe, however, that even the stronger, no-evidence claim might be defensible.

Fifth, on my account the PCS is not explicitly committed to the claim that ϕ is unique to phenomenal concepts.[5] It might be argued that (2) is true only if ϕ *is* unique to phenomenal concepts. For example, it might be argued that the epistemic gap between physical and phenomenal truths is unique, and so whatever property of phenomenal concepts explains the epistemic gap must be unique to phenomenal concepts. Proponents of certain versions of the PCS can accept this consequence. But they do not have to. They might argue, as many materialists have, that the epistemic gap between physical and phenomenal

materialist constraint.

[4]For example, Balog (2012b) concedes to anti-materialists that the relevant property of phenomenal concepts cannot be physically explained (p. 16), but she seems to argue that the fact that phenomenal concepts have this property is still no evidence of an ontological gap (pp. 16-21).

[5]Given this, my initial gloss of the PCS was slightly misleading. I said that proponents of the PCS claim that the epistemic gap arises because of something *special* about phenomenal concepts. Really they claim merely that the epistemic gap arises because of *something* about phenomenal concepts, whether or not this something is special.

truths is *not* unique.[6]

2.3 Why the conclusion of the PCS matters

On my account, the PCS aims to show that the epistemic gap is no evidence of an ontological gap. If this is right, then the conceivability of zombies is no evidence of their possibility, and Mary's ignorance of what it is like to see red is no evidence that phenomenal redness is ungrounded in physical properties. Although this conclusion is consistent with the soundness of the epistemic arguments (for reasons I discuss below), it is still extremely significant. It completely undermines their force. If the epistemic gap is no evidence of an ontological gap, then the epistemic arguments are *bad* even though they might still be sound.

To see the this, it might help to consider an analogy. Consider the following argument:

(1) Sally has spots

(2) If Sally has spots, then she has the flu.

(3) ∴ Sally has the flu.

This is a bad argument because the fact that someone has spots is no evidence that they have the flu. The argument might, however, still be sound. It might be that Sally has spots and also, by chance, has the flu. Then the material conditional 'if Sally has spots, then she has the flu' is true. But unless we have

[6]Some notable examples: Block and Stalnaker (1999) and Byrne (1999) argue that there are epistemic gaps between the physical truths and truths about ordinary macroscopic objects. Schaffer (2017) argues that epistemic gaps are even more widespread than this. For example, he argues that there is an epistemic gap between the truth that there are binded H, H, and O atoms and the truth than there is an H_2O molecule. Levine (1998, 2000) and Papineau (2011) also seem sympathetic to the view that the epistemic gap between physical and phenomenal truths is not unique. See §2.6.1 for more on this point.

some independent reason to believe that Sally has the flu (or that she does not have spots), we have no reason to believe that if Sally has spots then she has the flu.

Proponents of the PCS say the same thing about the epistemic arguments. They say that they are bad arguments because the epistemic gap is no evidence of an ontological gap. Like before, the arguments might still be sound. It might be that there is an epistemic gap and also, by chance, an ontological gap. But unless we have some independent reason to believe that there is an ontological gap (or that there is no epistemic gap), we have no reason to believe (E→O).

It would be nice if the PCS could show that (E→O) is false and the epistemic arguments are unsound. But doing that would be a tall order. (E→O) is a material conditional. To show that it is false, you would have to show that its antecedent is true and its consequent is false. And its consequent is the negation of materialism. So to show that (E→O) is false, you would have to show that materialism is true. But the PCS was never supposed to be a positive argument for materialism. It was just supposed to show that the epistemic arguments fail.

Many take the PCS to be arguing not that the epistemic gap is no evidence of an ontological gap, but rather that the epistemic gap can be physically explained.[7] Balog (2012b) takes the PCS to be arguing for an even weaker conclusion: that the epistemic gap can be explained in a way that is *compatible* with materialism. It is not clear how these conclusions pose a problem for anti-materialist arguments. Perhaps they are supposed to entail that the epistemic gap is no evidence of an ontological gap, though these entailments are not obvious.

These other conclusions are not altogether without interest. For example, if the epistemic gap can be physically explained, then that plausibly does *something* to weaken the force of the epistemic arguments. Still, it is far from obvious just

[7]For example, see Chalmers (2007) and Levine (2007).

how damaging these conclusions are to them. Chalmers (2007) suggests that even if the epistemic gap can be physically explained, "one might hold that the residual first-order explanatory gap still poses a problem for physicalism" (p. 172). So I think that it is better to present the PCS as I have, as arguing directly that the epistemic gap is no evidence of an ontological gap.

2.4 Versions of the PCS

Up to this point I have been concerned with the general structure of the PCS. To get a specific version of the PCS, we must substitute some property for ϕ in the above argument schema. Here I will focus on five main versions of the PCS, each of which substitutes a different property for ϕ, and consider their prospects. I am most sympathetic to the first main version I consider—the analyzability version—together with the closely related recognitional and quotational versions. The rest, I will argue, have serious problems.

2.4.1 The analyzability version

The analyzability version of the PCS claims that phenomenal concepts are non-analyzable into physical or functional concepts (simply 'non-analyzable' for short[8]). Here conceptual analysis is understood in terms of the a priority of certain sentences.[9] Roughly, the concept expressed by the predicate *F* is ana-

[8]Do not be mislead by this shorthand. "Non-analyzable" concepts, in this sense, might still be analyzable into other concepts. They just cannot be analyzed into *physical or functional* concepts. For example, <being in pain and experiencing phenomenal redness> might count as non-analyzable in this sense, even though it can be analyzed into the concepts <being in pain> and <experiencing phenomenal redness>.

[9]Following Chalmers (2012), I am working with a a understanding of a priority on which it is a property of sentences, not propositions. A sentence is *a priori* iff anyone who understands it is in a position to know that it is true. This avoids some complications that arise if a priority is understood, as it is more commonly, as a property of propositions. See Chalmers (2012, pp. 185-8)

lyzable into the concepts expressed by the predicates $G_1 \ldots G_n$ iff a sentence of the form 'for all x, Fx iff _____' is a priori, where the blank is filled in only by logical expressions and the predicates $G_1 \ldots G_n$.[10] So the concept <bachelor> is analyzable into the concepts <unmarried> and <male> because the sentence 'for all x, x is a bachelor iff x is an unmarried male' is a priori.

The analyzability version of the PCS can be formalized as follows:[11]

(A1) Phenomenal concepts are non-analyzable.

(A2) The fact that phenomenal concepts are non-analyzable explains the epistemic gap.

(A3) The fact that phenomenal concepts are non-analyzable is no evidence of an ontological gap.

(A4) ∴ The epistemic gap is no evidence of an ontological gap.[12]

(A1) should be relatively uncontroversial. The only people who deny (A1) are analytic functionalists, who claim that phenomenal concepts *can* be analyzed into physical or functional concepts. For example, analytic functionalists might claim that <pain> might be analyzed as <that which tends to cause distress and which in turn tends to be caused by bodily damage>. These analyses are now generally rejected. Of course, it might be true that pain *is* some functional

for discussion.

[10] Alter (2015) also glosses analyzability in terms of the a priority of certain sentences of this form. Two notes: (i) This gloss is not quite right, since junk can added to the sentence. For example, 'for all x, x is a bachelor iff (x is an unmarried male and 2+2=4)' is also a priori, but <bachelor> cannot be analyzed into mathematical concepts. (ii) This gloss covers only concepts of ordinary properties and not concepts of individuals or logical concepts. Some other gloss would be needed to cover them.

[11] Levine (2014) seems sympathetic to this version of the PCS, though he does not present it as I do. He also puts things in terms of semantic primitiveness instead of non-analyzability.

[12] Throughout this section I omit the bridge premise, which is common to all versions of the PCS. See §2.5 for a defense of the bridge premise.

state. But it is now generally denied that it is part of the *concept* of pain that it a functional state.[13]

In any case, analytic functionalism should be resisted by proponents of the epistemic arguments. This is because these anti-materialists must accept (E), the claim that there is an epistemic gap between the physical and phenomenal truths. But if analytic functionalism is true, then (E) is false: knowing all the physical truths puts you in a position to know what functional states things are in, and so if analytic functionalism is true then this also puts you in a position to know what phenomenal states they are in.[14] So anti-materialists can deny (A1) only by also denying a key premise of the epistemic arguments.

That just leaves (A2) and (A3). The quick argument for (A3) is to note that we have some non-analyzable concepts of uncontroversially physical things. Consider, for example, the concept <life>. As Block and Stalnaker (1999, pp. 14-15) suggest, there is arguably no analysis of <life> into physical or functional concepts. But this is no evidence that every living thing is non-physical or that there is an ontological gap between the physical truths and truths about living things. Given this, it seems plausible that the mere fact that *phenomenal* concepts are non-analyzable is no evidence of an ontological gap between physical and phenomenal truths.

(A2) can be motivated by a general principle about analyzability and epistemic gaps. This principle says, roughly, that conceptual analysis is necessary to

[13]See Lewis (1966, 1972) for a defense of analytic functionalism and Block (1978) for some classic arguments against functionalism. See Levin (2013) for a comprehensive overview of both analytic and non-analytic functionalism.

[14]Two qualifications: (i) In principle, you could think that there is an epistemic gap between the physical truths and ordinary functional truths, such as truths about what functional state my body or brain is in. But if this is right, then it is false that epistemic gaps in general entail ontological gaps (since there is no *ontological gap* between the physical truths and truths about the functional state of my brain or body). This, in turn, undermines (E→O) for reasons I explain in §4.3.2. (ii) If materialism is false, then there will be at least *some* functional truths that knowing the physical truths does not put you in a position to know, namely those involving the non-physical phenomenal states.

close epistemic gaps. More precisely:

> **Analysis Principle:** If the concepts involved in one truth differ from and cannot be analyzed into the concepts involved in another truth, then there is an epistemic gap between the two truths.

If the Analysis Principle is true, then the fact that phenomenal concepts are non-analyzable entails that there is an epistemic gap between physical and phenomenal truths. If this is right, then the fact that phenomenal concepts are non-analyzable arguably *explains* the epistemic gap and (A2) is true.[15]

The key question, then, is whether the Analysis Principle is true. The principle does have some intuitive appeal (how *could* you infer one truth from another if there is no logical or conceptual connection between them?), and various philosophers have endorsed it.[16] Still, it is far from obvious. On top of this, it has been strongly criticized by Chalmers and Jackson (2001) and Chalmers (2014), who give various putative counterexamples to it. So this is the weak-link of the analyzability version of the PCS. One option for materialists is to try to defend the Analysis Principle against Chalmers and Jackson's objections. But I think a better option is to simply conclude more modestly that *if* the Analysis Principle is true, *then* this version of the PCS succeeds.[17]

[15] I am not assuming that, in general, if p entails q, then p explains q. This general principle is false: that snow is white entails that it is either raining or not raining, but it does not explain why it is either is raining or not raining. Nonetheless, entailment is often closely tied to explanation, and in ordinary cases the fact that p entails q gives us some (defeasible) reason to think that p also explains q.

[16] Levine (1998), Balog (1999), Tye (2001) all seem to accept the Analysis Principle, though none do so explicitly. Even Chalmers seems to accept it in his earlier work. For example, here is what he says in his (1996) book, in arguing that there is an epistemic gap between physical and phenomenal truths:

> For consciousness to be [a priori] entailed by a set of physical facts, one would need some kind of analysis of the notion of consciousness... and there is no such analysis to be had (p. 104).

In later work, as I note above, he explicitly rejects the Analysis Principle.

[17] See chapter 5 for an extended discussion of the Analysis Principle.

In chapter 4 I explain in detail how even this modest conclusion helps undermine the epistemic arguments, but very briefly the idea is this. Many anti-materialists themselves rely on the Analysis Principle in defending (E), the claim that there is an epistemic gap.[18] In fact, I argue, the *only* plausible defense of (E) relies on the Analysis Principle. If this is right, then materialists do not need to take a stand on the truth or falsity of the Analysis Principle. If it is true, then the analyzability version of the PCS succeeds and (E→O) is unjustified. If it is false, then the only plausible defense of (E) fails and (E) is unjustified. Either way, materialists can plausibly reject at least one premise of the epistemic arguments. So even if the analyzability version of the PCS fails, it can play an important role in a defense of materialism.

Note on two related versions of the PCS: The analyzability version of the PCS is closely related to the recognitional and quotational versions of the PCS. According to the *recognitional* version, phenomenal concepts are recognitional concepts—concepts that you grasp in virtue of possessing a recognitional ability, and not in virtue of associating some description with them (see, for example, Loar 1990/7, Tye 2000, and Levin 2007). According to the *quotational* version, phenomenal concepts are quotational concepts—concepts that refer to the mental states they are constituted by, in a way analogous to quote terms[19] (see, for example, Papineau 2002, Block 2006, and Balogb 2012).

Proponents of both these versions of the PCS generally claim that recognitional or quotational concepts are non-analyzable into physical or functional concepts, and that this is ultimately what explains the epistemic gap.[20] As we

[18] See, for example, the Chalmers quote in footnote 16.

[19] A quote term is what you get when you put quotation marks around an expression to get a name of the expression that is inside the quotation marks. (For example, "dog" is a quote term that refers to the expression 'dog'.) The quotational version of the PCS is defended by Papineau (2002, 2007), Block (2006), and Balog (2012b). Chalmers (2003b) is sympathetic to the quotational account of phenomenal concepts, but he does not attempt to use it to defend the PCS.

[20] Papineau is an exception. He claims that phenomenal concepts are quotational, but gives a

have seen, this is precisely what proponents of the analyzability version claim about phenomenal concepts. Given this, I think that it is better for these proponents of the PCS to claim simply that phenomenal concepts are non-analyzable, as proponents of the analyzability version do, without taking a stand on the exact nature of phenomenal concepts. This allows them to side-step debates over the exact nature of phenomenal concepts.

2.4.2 The indexical version

The indexical version of the PCS claims that phenomenal concepts are indexical concepts—concepts like <here>, <now>, and <I>. It is well-known that there is an epistemic gap between the physical truths and truths involving indexical concepts. For example, knowing all of the physical truths does not put me in a position to know where I am or what time it is. So if phenomenal concepts are indexicals, then this might explain why there is an epistemic gap between physical and phenomenal truths. Perry (2001) is the most notable defender of this version of the PCS.

The indexical version of the PCS can be formalized as follows:

(I1) Phenomenal concepts are indexicals.

(I2) The fact that phenomenal concepts are indexicals explains the epistemic gap.

(I3) The fact that phenomenal concepts are indexicals is no evidence of an ontological gap.

(I4) ∴ The epistemic gap is no evidence of an ontological gap.

different explanation of the epistemic gap. See §2.4.4 below.

(I3) should be uncontroversial. The mere fact that we have indexical concepts is no evidence of an ontological gap. No one would try to argue against materialism by pointing to the fact that we have indexical concepts like <here> or <now>. So the crucial premises are (I1) and (I2). The problem is that there is good reason to reject both of them.

Against (I1), Chalmers (2004) argues that phenomenal concepts are not indexicals. His idea is that ignorance of indexical truths "disappears" from the third-person perspective (in a sense explained in the next paragraph), but ignorance of phenomenal truths does not disappear from the third-person perspective. Here an indexical truth is any truth that involves an indexical concept, and a phenomenal truth is any truth that involves a phenomenal concept. If this disappearance claim is right, then it follows that phenomenal concepts are not indexicals.

Chalmers explains the disappearance-of-ignorance idea with the following example. Suppose that there are two identical twins, A and B, one in America and one in Australia. It might be that that each twin is ignorant of where he is, even if each knows that A is in America and B is in Australia (suppose that each is also ignorant of whether he is A or B). But from the third-person perspective, we can have no corresponding ignorance. In knowing where A is and where B is, we know everything there is for us to know about the locations of the twins.[21]

But, Chalmers points out, the same is not true of phenomenal knowledge. For example, think about Mary in the black-and-white room. She is ignorant of what it is like to see red. But even from the third-person perspective, there *is* something here for us to know or be ignorant of—namely, what it is like to see red. The ignorance is not necessarily unique to Mary. So, unlike ignorance

[21]This example is presumably inspired by Lewis' (1979) twin god example, which illustrates what Lewis calls 'de se' ignorance.

of indexical truths, ignorance of phenomenal truths does not disappear from the third-person perspective. Given this, (I1) is false: if phenomenal concepts were indexicals, then phenomenal knowledge would be indexical and so *would* disappear from the third-person perspective.

(I2) does not fare any better. Against (I2), it is not obvious that the fact that phenomenal concepts are indexicals *would* explain the epistemic gap between physical and phenomenal truths. It is true, as I mentioned above, that indexical concepts do generate (and explain) a certain sort of epistemic gap: knowing all the physical truths does not put me in a position to know where I am or what time it is, for example. But it is not obvious that this is the sort of epistemic gap that we are faced with between physical and phenomenal truths.

To see this, think about indexical ignorance of where I am. If I know all the physical truths, then I know the range of possible locations that I could occupy. I just do not know which specific location I occupy. To tell me where I am, you just need to point to a particular location which I already know about and say that that is the location I occupy. You do not need to give me any new information about what the world is like.

But when it comes to ignorance of phenomenal truths, something different is going on. Even if I know all the physical truths, I still do not know the range of possible experiences in the world. It is not as if I know that there is pain, pleasure, phenomenal redness, and so on, but I am ignorant merely of my location in the world with respect to those experiences. Knowing all the physical truths does not put me in a position to know *anything at all* about pain, pleasure, or phenomenal redness.[22] I am not missing something merely about my location in the world. I am missing something about the world itself.

[22] At least, this is the intuition Jackson appeals to when making the knowledge argument, and proponents of the PCS are willing to grant him this.

So maybe the fact that phenomenal concepts are indexical concepts would be able to explain why there is some kind of epistemic gap between the physical and phenomenal truths. But it would be able to explain only the sort of epistemic gap that can be closed by the mere addition of locating information, like the epistemic gap between the physical truths and the truth about where I am. Since the epistemic gap between physical and phenomenal truths *cannot* be closed by the mere addition of locating information, indexicality alone cannot explain this epistemic gap. Given this, (I2) is false.

2.4.3 The possession version

The possession version of the PCS claims that phenomenal concepts can be possessed only by those who have experienced the relevant phenomenal states. For example, perhaps the concept of phenomenal redness can be possessed only by those who have experienced phenomenal redness. This version of the PCS can be formalized as follows:[23]

(P1) Phenomenal concepts can be possessed only by those who have experienced the relevant phenomenal states.

(P2) The fact that phenomenal concepts can be possessed only by those who have experienced the relevant phenomenal states explains the epistemic gap.

(P3) The fact that phenomenal concepts can be possessed only by those who have experienced the relevant phenomenal states is no evidence of an ontological gap.

[23]Stoljar (2005) criticizes the possession version of the PCS. He cites Tye (1999) and Papineau (1999) as defenders of it, though I am unsure whether Papineau (1999) actually defends it. Papineau (1999) does claim that phenomenal concepts can be possessed only by those who have experienced the relevant phenomenal states, but as far as I can tell he does not use this claim to defend the PCS.

(P4) ∴ The epistemic gap is no evidence of an ontological gap.

(P1) has been endorsed by many philosophers, going back at least to the empiricists.[24] The main contemporary opposition to it comes from views according to which not much is required to possess a concept. For example, Burge (1979) famously argued that you can possess the concept <arthritis> even if you falsely believe that someone can have arthritis in their thigh. This might be because it is possible to possess a concept deferentially, where the content of the concept depends in part on how it is used in the broader community. Inspired by this sort of view, Ball (2009) argues that you can grasp the concept <phenomenal redness> even if you have never experienced phenomenal redness.

But even if Ball is right about this, there is an easy fix, suggested by Alter (2013): substitute talk about concept possession in (P1)-(P3) with talk about concept mastery. Even if it is possible to deferentially possess the concept <phenomenal redness> without having experienced phenomenal redness, it is plausible that it is not possible to *master* the concept <phenomenal redness> without having experienced phenomenal redness. So a revised (P1), which talks about conceptual mastery, avoids Ball's objection. I think that this revised version of (P1) is plausible.

(P3) is also plausible. The conditions for our possessing or mastering a certain type of concept seem to be just a psychological fact about us. This is no evidence for any conclusion about the fundamental nature of the universe. In particular, it is no evidence of an ontological gap between the physical and phenomenal truths.

The problem is with (P2). It is not obvious how facts about concept possession

[24]The empiricists endorsed the more general claim that *any* concept can be possessed only by those who have had the relevant experiences. For example, Hume's *Enquiry* (1748/2011) says that "all our ideas [i.e., concepts]...are copies of our impressions" (§II), and so you cannot poses a concept without having had the relevant "impression" or experience.

or mastery are supposed to explain the all of the relevant epistemic gaps. The easiest way to see this is to think about conceivability gaps, such as the fact that zombies are conceivable. Nothing about the possession conditions for phenomenal concepts explains the conceivability of zombies. After all, zombies are conceivable only for those who possess the relevant phenomenal concepts. So even if the possession conditions for phenomenal concepts are unusually strict, that does not help explain why those who *do* possess the relevant concepts *can* conceive of zombies.

Proponents of (P2) are presumably thinking of knowledge gaps and not conceivability gaps. For example, maybe the possession conditions for phenomenal concepts can explain why Mary, while in the black-and-white room, cannot learn what it is like to see red. If the concept <phenomenal redness> can be possessed only by those who have experienced phenomenal redness, then Mary does not possess it. And if you do not possess a concept, then you cannot entertain any thoughts that involve it. Given this, Mary would not be able to infer the truth about what it is like to see red from the physical truths, even if they did entail this truth.

But even if this is right, it does not work as a general response to the knowledge argument. There are versions of the knowledge argument on which Mary possesses the relevant concepts, but still cannot learn certain phenomenal truths. For example, Stoljar (2005, p. 486-7) has us imagine that Mary is allowed to see red and gains the concept <phenomenal redness>, but then suffers amnesia after she returns to the black-and-white room. She retains the concept <phenomenal redness> and can still imagine phenomenal redness, but she cannot remember which physical things (say, tomatoes) produce it. But even though she possesses the concept <phenomenal redness>, she still cannot infer the truth about which

things cause phenomenal redness from the physical truths.[25]

So maybe facts about concept possession can explain the specific epistemic gap that Mary faces in Jackson's original thought experiment.[26] But facts about concept possession cannot explain the conceivability gap or even certain knowledge gaps, such as the one faced by Stoljar's Mary. And if proponents of the PCS want to say that *no* anti-materialist argument succeeds, then they must explain *every* epistemic gap. If they can explain the knowledge gap but not the conceivability gap, then the conceivability argument might succeed even if the knowledge argument fails. And if they can explain one type of knowledge gap but not others, then some versions of the knowledge argument might succeed even if one version of it fails.

2.4.4 The antipathetic fallacy version

Papineau (1993, 2002) is a proponent of the quotational account of phenomenal concepts, which I discussed in §2.4.1 above. But although quotational concepts are non-analyzable, Papineau does not point to this property of them in trying to explain the epistemic gap. Instead, he tries to explain the epistemic gap by claiming that quotational concepts give rise to what he calls 'the antipathetic fallacy' (2002, pp. 169-71).[27]

The basic idea is this. Since phenomenal concepts are quotational, you cannot entertain a phenomenal truth that involves such a concept without being in the relevant phenomenal state.[28] For example, I cannot think that I am in pain by

[25]Note that in this case, unlike in Jackson's original thought experiment, Mary's ignorance is not of what it's like to see red. See Nida-Rümelin (1995) for a similar variation on Jackson's thought experiment.

[26]Of course, even this is doubtful, given that facts about concept possession cannot explain the epistemic gap faced by Stoljar's Mary, or the conceivability gap. Since these three epistemic gaps are relevantly similar, a good explanation of *any* of them should arguably explain *all* of them.

[27]Tye (1999, pp. 71-3) offers a similar explanation of the epistemic gap.

[28]At least, this is true on a simple account of quotational concepts. On slightly more sophis-

using the quotational concept <pain> without being in pain. Now suppose that pain is identical to c-fiber firing. <c-fiber firing> is not a quotational concept. So when I think that I am in pain by using the concept <c-fiber firing>, I do not need to be in pain. Papineau thinks this difference leads us to believe that pain cannot be identical to c-fiber firing. So we tacitly make something like the following argument:

(1) I necessarily feel pain when I use the quotational concept <pain>.

(2) I do not necessarily feel pain when I use the concept <c-fiber firing>.

(3) $\therefore$ Pain $\neq$ c-fiber firing.

The argument is obviously invalid. You commit the antipathetic fallacy if you take this form of argument to be valid.

This version of the PCS can then be formalized as follows:

(F1) Phenomenal concepts give rise to the antipathetic fallacy.

(F2) The fact that phenomenal concepts give rise to the antipathetic fallacy explains the epistemic gap.

(F3) The fact that phenomenal concepts give rise to the antipathetic fallacy is no evidence of an ontological gap.

(F4) $\therefore$ The epistemic gap is no evidence of an ontological gap.

(F3) is plausible. The fact that a concept makes us disposed to commit a certain fallacy is no evidence of an ontological gap between the physical truths and truths involving that concept. The trouble is with (F1) and (F2).

---

ticated accounts (for example, the account given in Papineau 2002), you do not have to be in the relevant phenomenal state, but you *do* have to at least be *imagining* the relevant phenomenal state.

To start, it is not clear that phenomenal concepts *do* give rise to the antipathetic fallacy, as (F1) says. For example, consider the above argument for the claim that pain $\neq$ c-fiber firing. That argument is so obviously invalid that it is hard to believe that anyone would be even subconsciously persuaded by it. Is every philosopher who thinks that there is an epistemic gap between physical and phenomenal truths really so confused? So on the face of it, I think that (F1) is implausible. Maybe there is some cognitive defect in us that makes us vulnerable to such a bad argument, but Papineau says nothing to show that there is.

Levine (2007, p. 163 note 8) makes a similar point. He notes that when you refer to something (e.g., your laptop) using the demonstrative 'that', the thing that you refer to is necessarily present. But when you refer to something using a description (e.g., 'my laptop'), the thing that you refer to is not necessarily present. This is analogous to how you are necessarily in pain when you use the quotational concept <pain> but you are not necessarily in pain when you use the concept <c-fiber firing>. Nonetheless, you are not disposed to think that *that* (pointing at your laptop) cannot be your laptop. So it is hard to see why you would be disposed to make the analogous mistake for pain and c-fiber firing.[29]

Worse, (F2) is false. To see this, consider the fact that people are disposed to make all sorts of bad arguments, but this does not entail anything interesting about epistemic gaps. For example, people are disposed to make bad arguments that aim to show that $0.999\ldots \neq 1$. This does not entail that there is an epistemic gap between the principles of mathematics and the truth that $0.999\ldots = 1$. There is, in fact, *no* such epistemic gap, and people who think more carefully are able to come to know that $0.999\ldots = 1$.

[29]Sundström (2008) raises a different objection to (F1). He points out that we are not disposed to doubt identities such as "TF's favorite color experience = phenomenal redness", even though the left side of the identity statement does not involve any quotational concepts and the right side does. But if (F1) is true, then we *should* be disposed to deny such identities. Papineau (2011, p. 17) briefly responds to this objection by denying that (F1) has this consequence.

Similarly, we might be disposed to make the above bad argument for the claim that pain $\neq$ c-fiber firing, and then conclude on this basis that there must be an epistemic gap between physical and phenomenal truths. But this does not mean that there is actually an epistemic gap between them. On more careful reflection, we might be able to infer phenomenal truths from physical truths, just as on more careful reflection we are able to infer the truth that $0.999\ldots = 1$ from the principles of mathematics.

So the antipathetic fallacy would not explain why there is an epistemic gap. Still, it might be able to explain why we would *believe* that there is an epistemic gap even if there were no epistemic gap, just as the fact that we are disposed to make bad mathematical arguments can explain why some of us *believe* that $0.999\ldots = 1$ does not follow from the principles of mathematics even though it does. Such an explanation would not help proponents of the PCS, who admit that there is an epistemic gap. But this could help materialists who deny that there is an epistemic gap, but try to explain away why we believe that there is one.[30]

I doubt that even this is useful to materialists, since I doubt that (F1) is true. But if there is some way to defend (F1), then this might be a promising strategy for materialists to pursue.

2.4.5 The dual conceptual roles version

The final version of the PCS that I will consider is proposed by Nagel (1974) in a footnote to his paper 'What is it like to be a bat?'.[31] This version was later developed more fully by Hill (1997) and Hill and McLaughlin (1999). But for simplicity I will follow Nagel's presentation of it here.

[30]This, in fact, is how Papineau responds to the epistemic arguments in a later (1998) paper.
[31]The label 'dual conceptual roles' comes from Chalmers (2007).

The basic idea is this. You can imagine something either perceptually or sympathetically. To imagine something perceptually is to "put [yourself] in a conscious state resembling the state [you] would be in if [you] perceived it". To imagine something sympathetically is to "put [yourself] in a conscious state resembling the thing itself" (Nagel 1974, p. 446, note 11). Now when we imagine zombies, we are combining our perceptual and sympathetic modes of imagination: we are perceptually imagining creatures physically just like us, and sympathetically imagining that they are not conscious.

Proponents of this version of the PCS then claim that although simply perceptually imagining that p or simply sympathetically imagining that p might be evidence that p is possible, the *combined* perceptual and sympathetic imagination that p is no evidence that p is possible. This is because, they claim, our imagination is not a reliable guide to possibility when its two different modes are combined in this way. Since our imagination of zombies is a combined perceptual and sympathetic imagination of them, it is no evidence that they are possible.

This version of the PCS does not quite fit the mold I laid out above, since it is does not claim simply that phenomenal concepts have a given property and that that is what explains the epistemic gap. Still, it follows largely the same pattern as before:

(D1) We can combine our perceptual and sympathetic modes of imagination to imagine zombies.

(D2) The fact that we can combine our perceptual and sympathetic modes of imagination to imagine zombies explains the epistemic gap.

(D3) The fact that we can combine our perceptual and sympathetic modes of imagination to imagine zombies is no evidence of an ontological gap.

(D4) $\therefore$ The epistemic gap is no evidence of an ontological gap.

Doggett and Stoljar (2010) discuss this version of the PCS at length, including some variations on it, so here I will be brief. Each of its premises can be challenged.

Against (D1), Doggett and Stoljar (2010, p. 131) object that we are not using our sympathetic imagination when we imagine zombies, and a fortiori are not combining it with our perceptual imagination. Again, to sympathetically imagine something is to "put [yourself] in a conscious state resembling the thing itself". So to sympathetically imagine a zombie is to put yourself in a conscious state resembling the conscious state of a zombie—"a conscious state resembling the state of having no conscious states", as Doggett and Stoljar put it. This is not obviously possible, but even if it is possible it is clearly not what we doing when we imagine zombies.

(D2) is also a problem. It is not obvious how our combining our perceptual and sympathetic modes of imagination is supposed to explain all of the relevant epistemic gaps. The easiest way to see this is to think about knowledge gaps, such as Mary's inability to learn, while in the black-and-white room, what it is like to see red. Mary's ignorance is not due to her combining her perceptual and sympathetic modes of imagination in an unreliable way. After all, she cannot perceptually imagine red or sympathetically imagine phenomenal redness even individually, let alone combine the imagination of red or phenomenal redness with anything else. If anything, the problem is that her imagination too limited, not that it is too wide-ranging.

That just leaves (D3). It is not obviously false, but even it is open to a serious challenge. The idea behind (D3) is that our combining our perceptual and sympathetic modes of imagination is an unreliable guide to possibility. The problem here is that combining different modes of imagination is not, in general,

an unreliable guide to possibility. For example, I can visually imagine an apple and olfactorily imagine the smell of an orange. Even though this combines different modes of imagination, it still seems to be evidence that it is possible for an apple to smell like an orange.[32] So (D3) is true only if there is something special about the perceptual and sympathetic modes of imagination that makes combining *them* an unreliable guide to possibility. A proponent of (D3) can be challenged to say what this is.

2.5 Evidence and explanation

Recall that my account of the PCS has the following form:

(1) *Attribution:* Phenomenal concepts are ϕ.

(2) *Explanation:* The fact that phenomenal concepts are ϕ explains the epistemic gap.

(3) *No evidence:* The fact that phenomenal concepts are ϕ is no evidence of an ontological gap.

(4) *Bridge:* If (1)-(3) are true, then (5) is true.

(5) $\therefore$ The epistemic gap is no evidence of an ontological gap.

In the previous section I considered whether (1)-(3) are plausible on various substitutions for ϕ. I will now consider (4), the bridge premise, which is common to all versions of the PCS.

The most natural way to defend the bridge premise is by appeal to general principles governing evidence and explanation. Here is one initially promising principle that does not quite work:

[32]Doggett and Stoljar (2010) make a similar point. Note that I am distinguishing modes of imagination more fine-grainly than Nagel does in the footnote, since he would classify both visual and smell modes of imagination as perceptual imagination.

(exp_1) If e explains p_1, and e is no evidence that p_2, then p_1 is no evidence that p_2.

Here 'e', 'p_1', and 'p_2' are schematic letters for propositions. (exp_1) straightforwardly entails the bridge premise.[33]

Unfortunately, (exp_1) is false because there can be misleading evidence. For example, suppose that there is a bloody knife in my apartment because someone planted it there, but that you do not know that the knife was planted. Then the fact that there is a bloody knife in my apartment (= p_1) is evidence for you that I am the murderer (= p_2). But the fact that there is a bloody knife in my apartment is explained by the fact that someone planted it there (= e), which is no evidence that I am the murderer. So (exp_1) entails incorrectly that the fact that there is a bloody knife in my apartment is no evidence that I am the murderer.

To get around this problem, we can weaken (exp_1) slightly to get (exp_2):

(exp_2) If e explains p_1 and that e is no evidence that p_2, then p_1 is *either* no evidence that p_2 *or* misleading evidence that p_2.

This modified principle entails correctly that the bloody knife is either no evidence that I am the murderer or misleading evidence that I am the murderer. (exp_2) does not quite entail the bridge premise, but it does entail the slightly weaker premise (4*):

(4*) *Weaker bridge:* If (1)-(3) are true, then the epistemic gap is either no evidence of an ontological gap or misleading evidence of an ontological gap.

This creates a small but harmless wrinkle for proponents of the PCS. Given the weaker premise (4*), the PCS is valid only if its conclusion is weakened to say

[33] To see this, substitute 'that phenomenal concepts are ϕ' for 'e', 'that there is an epistemic gap' for 'p_1', and 'that there is an ontological gap' for 'p_2' to get (making minor changes for grammar and readability): if the fact that phenomenal concepts are ϕ explains the epistemic gap (= (2)), and the fact that phenomenal concepts are ϕ is no evidence of an ontological gap (= (3)), then the epistemic gap is no evidence of an ontological gap (= (5)).

that the epistemic gap is *either* no evidence of an ontological gap *or* misleading evidence of an ontological gap. But even this conclusion is strong enough to undermine the epistemic arguments. If materialists can persuade you that it is true, then you should reject the epistemic arguments. Compare: if I can persuade the jury that the bloody knife is either no evidence or misleading evidence that I am the murderer, then I should be acquitted.

Another complication arises from the fact that explanations often take for granted the presence of certain background conditions that are left unstated in the explanations themselves. For example, the spark explains why there was a fire, given the background conditions of oxygen and fuel. Since these background conditions are not stated in the explanation of the fire, we get a counterexample to (exp_2): the spark, by itself, is no evidence that there was fuel, the spark explains the fire, but (contrary to what (exp_2) entails) the fire *is* non-misleading evidence that there was fuel. After all, if I learn that there was a fire, I can conclude on that basis that there was fuel.

To get around this problem, I stipulate that 'explanation' in (exp_2) is to be understood as 'explanation-together-with-background-conditions'. On this understanding of 'explanation', the spark alone does not explain the fire, but the spark together with the oxygen and fuel do explain the fire. This avoids the previous counterexample to (exp_2) because the spark together with the oxygen and fuel is trivially evidence that there was fuel.

This creates another small wrinkle for proponents of the PCS. If the fact that phenomenal concepts are ϕ explains the epistemic gap only together with certain background conditions, then these background conditions must be added to (2) and (3). The possible problem here is with (3). In principle, one of the background conditions could itself be evidence of an ontological gap, and so the

modified (3) could be false even if the fact that phenomenal concepts are ϕ is no evidence of an ontological gap. But in practice, most putative explanations of the epistemic gap have either no background conditions or only background conditions that themselves are no evidence of an ontological gap.[34]

A final complication arises from the fact that on one conception of evidence, what something is evidence of can vary from person to person. For example, suppose that you do not know that spots are reliably correlated with measles. Then, on this conception of evidence, the fact that someone has spots is no evidence for you that they have measles. But if I know that spots are reliably correlated with measles, then the fact that someone has spots is evidence for me that they have measles.[35] Given this, (exp_2) cannot simply talk about whether, say, *e* is evidence that p_2. It must talk about whether *e* is evidence that p_2 *for a certain person*.

The quick fix is to simply add in the relevant qualifications to (exp_2) and to the PCS. This does not make (exp_2) or the PCS any less plausible. Alternatively, we can stipulate that the relevant conception of evidence is one on which what something is evidence of cannot vary from person to person. On this conception of evidence, the fact that someone has spots is evidence that they have measles, regardless of whether anyone knows that spots are reliably correlated with measles. This is what Kelly (2014) calls 'indicator evidence', since it is a reliable indicator of what it is evidence of. In any case, none of this poses a problem for the PCS.

[34]Distinguish between (i) the mere existence of an epistemic gap and (ii) our *facing* an epistemic gap (for example, noticing that one exists or making a fuss about it). It is plausible that non-analyzability *alone* explains (i) if the Analysis Principle is true. In contrast, non-analyzability explains (ii) only together with certain background conditions—say, our being able to form complex beliefs and having time to engage in philosophical reflection. But neither of these background conditions is evidence of an ontological gap.

[35]This example is from Kelly (2014).

2.6 Objections to the PCS

I will end this chapter by responding to some common objections to the PCS. I considered objections to specific versions of the PCS above, in §2.4. The objections that I will consider in this section are more general, and aim to show that most or all versions of the PCS fail. I will argue that none of these objections succeeds.

2.6.1 Too many epistemic gaps: non-phenomenal truths

The first objection to the PCS that I will consider targets only versions of the PCS that try to explain the epistemic gap by pointing to some property of phenomenal concepts that is not unique to them. This includes the analyzability, recognitional, and indexical versions of the PCS, since we have non-phenomenal concepts that are non-analyzable, recognitional, and indexical. The quotational and antipathetic fallacy versions of the PCS are also vulnerable to this objection if non-phenomenal concepts can be quotational.

The objection goes like this. Suppose that ϕ explains the epistemic gap between physical and phenomenal truths. If ϕ is not unique to phenomenal concepts, then there are epistemic gaps between physical truths and truths involving the other concepts that are ϕ. For example, suppose that ϕ is the property of being a recognitional concept. If the fact that phenomenal concepts are recognitional explains the epistemic gap between physical and phenomenal truths, then there are epistemic gaps between physical truths and truths involving other recognitional concepts. After all, if there were no such corresponding gaps, then the mere fact that phenomenal concepts are recognitional would not explain the epistemic gap between physical and phenomenal truths.

But, proponents of this objection go on to claim, there *not* any other such

epistemic gaps. For example, we have recognitional concepts of plants, but there are no epistemic gaps between physical truths and truths about plants. After all, if there were these epistemic gaps, then philosophers would have tried to use them to argue against materialism. But none have. Their focus has been exclusively on the epistemic gap between physical and *phenomenal* truths. This suggests that the epistemic gap between physical and phenomenal truths is unique. If this is right, then if ϕ is not unique to phenomenal concepts, then ϕ does not explain the epistemic gap between physical and phenomenal truths.

This objection can be formalized as follows:

(1) *Explanatory constraint:* If ϕ is not unique to phenomenal concepts, then ϕ explains the epistemic gap between physical and phenomenal truths only if there are many epistemic gaps.

(2) *Gap uniqueness:* There are not many epistemic gaps: the epistemic gap between physical and phenomenal truths is unique.

(3) $\therefore$ If ϕ is not unique to phenomenal concepts, then ϕ does not explain the epistemic gap between physical and phenomenal truths.

Again, this objection targets only versions of the PCS that try to explain the epistemic gap by pointing to a property of phenomenal concepts that is not unique to them. For the reasons that I gave above, I think that (1) is plausible. If there is no epistemic gap between the physical truths and truths involving the non-phenomenal concepts that are ϕ, then the mere fact that phenomenal concepts are ϕ cannot explain the epistemic gap between physical and phenomenal truths. So the best option for proponents of the PCS is to deny (2).[36]

[36] As I noted in chapter 1, Levine (2001, p. 50) distinguishes between what he calls 'type E' ("exceptionalist") materialists, who believe that the epistemic gap between physical and phenomenal truths is unique and 'type NE' ("non-exceptionalist") materialists, who believe that the epistemic gap between physical and phenomenal truths is not unique. So type E materialists accept (2), and type NE materialists reject it. Levine himself accepts type NE materialism.

Here the reasonable but flat-footed strategy for proponents of the PCS is to simply appeal to the same principle that they appealed to in explaining the epistemic gap to argue against (2). For example, consider the analyzability version of the PCS. Proponents of it appeal to the Analyzability Principle in explaining the epistemic gap. Recall that the Analyzability Principle says that if the concepts involved in one truth differ from and cannot be analyzed into the concepts involved in another truth, then there is an epistemic gap between the two truths. If this is true, then there is an epistemic gap between the physical truths and the truth that there are living creatures, since there is no analysis of <life> into physical concepts. This gives us a principled reason to think that the epistemic gap between physical and phenomenal truths is not unique.

It might be objected that this consequence gives us a reason to reject the Analysis Principle and other principles that have the same consequence, since it is obvious that the epistemic gap between physical and phenomenal truths is unique. In response, I deny that it is pre-theoretically obvious how many epistemic gaps there are. This is a hard question, and our intuitions are unlikely to be of much use in answering it (see chapter 4 for more on this point). On top of this, many philosophers have thought that the epistemic gap between physical and phenomenal truths is not unique, including Levine (1998, 2001), Block and Stalnaker (1999), Byrne (1999), and Schaffer (2017). Of course, they might all be wrong about this, but they are not *obviously* wrong. So the mere fact that the Analysis Principle entails that there are many epistemic gaps is not a good reason to reject it.

Why, then, have philosophers focused exclusively on the epistemic gap between physical and phenomenal truths, if there really are so many epistemic gaps? There are a couple of possible responses to this question.

One response is to deny that philosophers *have* focused exclusively on the epistemic gap between physical and phenomenal truths. For example, the British Emergentists thought that chemical truths are not fully grounded in physical truths. They motivated their view by claiming that there is an epistemic gap between physical and chemical truths.[37] Another example comes from Descartes' *Discourse on Method* (1637/1998), where he claims that it is inconceivable that something physical use language (§V).[38] If this is right, then there is an epistemic gap between physical truths and truths about language use.[39] It is only relatively recently in the history of philosophy that philosophers zeroed-in on the epistemic gap between physical and phenomenal truths.

Another response is to note that some things are obviously physical, whether or not there is an epistemic gap between the physical truths and truths about them. For example, suppose that there is an epistemic gap between the physical truths and certain truths about plants. We would still not expect philosophers to argue on this basis that materialism is false, since plants are obviously physical. This is because there are some simple heuristics that we can use to learn that plants are physical. For example, the following principle is plausible: if you can kick it, then it is physical. Since we can kick plants, we can use this heuristic to learn that plants are physical.[40] But we cannot obviously use this heuristic to learn that phenomenal states are physical, since phenomenal states cannot obviously be kicked. This might be why, although it is obvious that plants are

[37]See, for example, Broad (1925, p. 63), quoted in Stoljar (2006, p. 135).

[38]More precisely, he claims that it is inconceivable that a *machine* use language. But he presumably would have accepted the claim that it is inconceivable that something physical use language, since he thought the physical world was entirely mechanistic.

[39]Both of these examples are from Stoljar (2006, chapter 7). I am following his gloss of British Emergentism, which he notes can be understood in a few different ways.

[40]There are other possible heuristics we can use to learn this. For example, the following principle is plausible: if multiple people can see it, then it is physical. Again, we can use this heuristic to learn that plants are physical, since the same plant can be seen by different people. But we cannot obviously use it to learn that phenomenal states are physical, since a given phenomenal state cannot obviously be seen by different people.

physical, it is not obvious that phenomenal states are physical.

To be clear, I am claiming neither that the heuristic in the previous paragraph gives a correct sufficient condition for being physical nor that phenomenal states cannot be kicked. Maybe there could be non-physical things that can be kicked, and maybe phenomenal states can be kicked. My point is just that there is a *plausible* heuristic which *obviously* entails that plants are physical, but which does not obviously entail that phenomenal states are physical. Even if this plausible principle is false or if phenomenal states can somehow but non-obviously be kicked, this is enough to explain why it seems obvious that plants are physical, but it does not seem obvious that phenomenal states are physical.

It is also worth stressing that I am not claiming that the above heuristic closes the supposed epistemic gap between physical truths and truths about plants. I am assuming for the sake of argument that there *is* such an epistemic gap—that knowing all the physical truths does not put you in a position to know all the truths about plants, and that it is conceivable that the physical truths hold without all the truths about plants holding. I am just claiming that even if that is right, you can still use the "can you kick it?" heuristic to learn that plants are physical. The epistemic gap remains. The heuristic merely allows you to conclude that *despite the epistemic gap*, plants are physical.

2.6.2 Too many epistemic gaps: phenomenal truths

The previous objection claimed that the leading versions of the PCS fail because they entail that there are more epistemic gaps involving non-phenomenal truths than there actually are. Stoljar (2005, pp. 479-81) raises a related objection. He argues that the leading versions of the PCS fail because they entail that there are more epistemic gaps involving *phenomenal* truths than there actually are. He

focuses on the following conditional:

> If (N) *x* is a number, then (∼S) *x* is not a red sensation.

Since this conditional is a priori, there is no epistemic gap between (N) and (∼S): knowing (N) puts you in a position to know (∼S), and it is inconceivable that both (N) is true and (∼S) is false. But, Stoljar claims, the leading versions of the PCS entail that there *is* an epistemic gap between (N) and (∼S). If this is right, then the leading versions of the PCS fail.

For example, consider the analyzability version of the PCS. It appeals to the Analysis Principle in explaining the epistemic gap between physical and phenomeal truths. But the Analysis Principle says that whenever one truth involves a concept that differs from and cannot be analyzed into the concepts involved in a different truth, then there is an epistemic gap between the two truths. Since <red sensation> cannot be analyzed into <number> or any other concepts in (N), the Analysis Principle entails that there is an epistemic gap between (N) and (∼S). Or consider the recognitional version of the PCS. It can appeal either to the Analysis Principle or to a narrower principle about recognitional concepts: that whenever one truth involves a recognitional concept, then there is an epistemic gap between it and truths that do not involve recognitional concepts. Assuming that phenomenal concepts are recognitional, we again get the consequence that there is an epistemic gap between (N) and (∼S).

In response, I admit that there is no epistemic gap between (N) and (∼S). I also admit that many of the principles that proponents of the PCS appeal to do, *as they are typically stated*, have the consequence that there is an epistemic gap between (N) and (∼S). But I deny that this means that no version of the PCS can succeed, since the relevant principles can be formulated in more careful ways that do not have this consequence.

It will help to start by thinking more generally about the sort of principle that proponents of the PCS appeal to in explaining the epistemic gap. Call principles like the Analysis Principle 'barrier theses'.[41] Barrier theses claim that there is an epistemic gap between truths of one sort and truths of another sort. Proponents of the PCS accept certain barrier theses, but so do many other philosophers. For example, Hume (1738/2011) famously claimed that you cannot derive an 'ought' from an 'is'. Put in terms of epistemic gaps: there is an epistemic gap between normative and non-normative truths. Another example is Perry (1978, 2001), who claimed that there is an epistemic gap between truths that involve indexicals, such as <I> or <here>, and non-indexical truths. For example, simply knowing all the physical truths does not put you in a position to know who or where you are.[42]

These barrier theses are generally regarded as plausible. At the very least, they are not obviously false. But on their simplest formulations, they are vulnerable to counterexamples, just as the Analysis Principle is. For example, consider the normative truth 'no Martians ought to be shot'. Knowing the non-normative truth that there are no Martians puts me in a position to know that no Martians ought to be shot. So we get a counterexample to Hume's claim.[43] Or consider the indexical truth 'I am not in the bathroom'. Knowing the non-indexical truth that no one is in the bathroom puts me in a position to know that I am not in the bathroom. So we get a counterexample to Perry's claim.

Nonetheless, it is still plausible that some claims in the ballpark of the Hume's claim and Perry's claim are true. It would be a mistake to dismiss these claims simply because there are edge cases where, on their simplest formulations,

[41] This label is from Restall and Russell (2010), who apply it to principles about logical derivability specifically and not to principles about epistemic gaps more generally.

[42] As we saw above, some proponents of the PCS use this principle to try to explain the epistemic gap. But the principle has wide acceptance, even among those who reject the PCS.

[43] This is based loosely on an example from Prior (1960).

they give the wrong results. The counterexamples are a challenge to state the principles more carefully, not a reason to dismiss them altogether. The same goes for the Analysis Principle and other principles that proponents of the PCS appeal to in explaining the epistemic gap. Yes, there are edge cases where, on their simplest formulations, they give the wrong results. But there might still be principles in the ballpark of them that are true.

OK, but how exactly should the Analysis Principle and the other principles be modified in response to the relevant counterexamples? The quick fix for the Analysis Principle is to restrict it to atomic a posteriori truths:

> **Weakened Analysis Principle (WAP):** If an atomic a posteriori truth involves a concept that differs from and cannot be analyzed into the concepts involve in another truth, then there is an epistemic gap between the two truths.

A sentence is atomic iff it has no sentences as proper parts. So 'A&B' is non-atomic since it has the sentences 'A' and 'B' has proper parts, and '∼S' is non-atomic since it has the sentence 'S' as a proper part. WAP is not vulnerable to Stoljar's counterexample to the original Analysis Principle, since WAP does not entail that there is an epistemic gap between (N) and (∼S). This is because (∼S)—"x is not a red sensation"—is non-atomic, and so WAP does not apply to it. WAP may look ad hoc, but *every* barrier thesis needs to be modified somehow to avoid Stoljar-style counterexamples to it. Unless you are willing to reject every modified barrier thesis (including Hume's and Perry's) on the same grounds, you should be willing to take WAP seriously.

It might be wondered whether WAP is strong enough to explain all of the epistemic gaps between physical and phenomenal truths. WAP can explain why there is an epistemic gap between physical truths and atomic phenomenal truths

(for example, 'I am in pain', 'this is what it is like to see red'), but it cannot obviously explain the epistemic gap between physical truths and *non*-atomic phenomenal truths (for example, 'there is no phenomenal feel common to all unpleasant experiences').[44] Proponents of the epistemic arguments typically point to atomic phenomenal truths when defending the epistemic gap premise of their arguments, but in principle they could point to non-atomic phenomenal truths.

In response, I will again look ahead to chapter 4. There I argue that the only plausible defense of the claim that there even is an epistemic gap between physical and phenomenal truths relies on some version the Analysis Principle. Suppose, then, that WAP is the only plausible version of the Analysis Principle. Then the only epistemic gaps that proponents of the epistemic arguments can show exist are those that WAP entails exist—ones involving atomic truths, and possibly conjunctions of atomic truths. These are precisely the epistemic gaps that proponents of the PCS can use WAP to explain. If this is right, then WAP is strong enough to explain all of the epistemic gaps that we have reason to believe exist.

2.6.3 Overgeneralization

Stoljar (2005, pp. 484-5; cf. 2006, pp. 51-3) also argues that the PCS fails because it overgeneralizes: it entails not only that the epistemic arguments fail, but also that certain uncontroversially successful arguments fail. He focuses on the implications that the PCS has for Putnam's (1963) perfect actor argument against

[44]WAP might be able to explain *some* epistemic gaps involving non-atomic phenomenal truths, such as conjunctions. For example, if it can explain why p is inscrutable from physical truths, then it can also explain why the conjunction of p and q is inscrutable from physical truths. But no such explanation works for epistemic gaps involving other non-atomic phenomenal truths, such as the example given in the main text.

behaviorism, which is widely thought to succeed.

As Stoljar defines behaviorism, it is the view that phenomenal truths are grounded in behavioral truths. Call this 'grounding behaviorism'. Grounding behaviorism entails that my having in a headache is grounded in how I behave and how I am disposed to behave. The perfect actor argument starts by claiming that it is conceivable that there is a perfect actor who, when he is in pain, neither behaves as normal people do when they are in pain nor is disposed to behave as normal people do when they are in pain. If this is right, then there is an *epistemic gap* between behavioral and phenomenal truths. On Stoljar's account, proponents of this argument then infer that that there is an *ontological gap* between behavioral and phenomenal truths, and consequently that grounding behaviorism is false.

The perfect actor argument, as presented by Stoljar, can be formalized as follows:

(E) There is an epistemic gap between behavioral and phenomenal truths.

(E→O) If there is an epistemic gap between behavioral and phenomenal truths, then there is an ontological gap, and grounding behaviorism is false.

(O) ∴ Grounding behaviorism is false.

This argument should look familiar. It has the same structure as the epistemic arguments. The only difference is that 'grounding behaviorism' has been substituted for 'materialism' and 'behavioral truths' has been substituted for 'physical truths'.

Stoljar claims that if the PCS succeeds and undermines premise (E→O) of the epistemic arguments, then it also undermines premise (E→O) of the perfect actor argument. Recall that the PCS attempts to undermine (E→O) by showing that the epistemic gap between physical and phenomenal truths is no evidence

of an ontological gap between them. But the behavioral truths are a subset of the physical truths. So if the epistemic gap between physical truths, in general, and phenomenal truths is no evidence of an ontological gap between them, then a fortiori the epistemic gap between *behavioral* truths, in particular, and phenomenal truths is no evidence of an ontological gap between them. This, in turn, undermines ($\underline{\text{E}}\rightarrow\underline{\text{O}}$) in the same way that the corresponding conclusion of the PCS undermines (E→O).

To get to the conclusion that the PCS fails, we just need to add the assumption that the perfect actor argument succeeds. If the perfect actor argument succeeds but the PCS entails that it fails, then the PCS itself fails.

In response, I deny that the perfect actor argument, as presented by Stoljar, succeeds. It may be true that the perfect actor argument, as presented by Putnam in his 1963 paper, succeeds. But Putnam's version of the argument differs significantly from Stoljar's version.

Putnam's version of the perfect actor argument targets not grounding behaviorism, but *logical* behaviorism: the view that our concepts of consciousness can be analyzed into behavioral concepts. Putnam does claim that it is conceivable that there is a perfect actor. So he is committed to ($\underline{\text{E}}$). But he does not conclude from this that that grounding behaviorism is false. Instead, he concludes more modestly that logical behaviorism is false. And when he does this, he is not moving from an epistemic premise to an ontological conclusion. So Putnam is not committed to ($\underline{\text{E}}\rightarrow\underline{\text{O}}$). His argument is uncontroversial only because it is uncontroversial that a *conceptual* conclusion (that logical behaviorism is false) follows from the epistemic premise, ($\underline{\text{E}}$). Given this, Putnam's version of the argument is not undermined by the PCS.[45]

[45]It should be clear to anyone who reads Putnam (1963) that this, and not Stoljar's argument, is the argument that Putnam makes. Throughout the article he is explicitly concerned with logical behaviorism, and never even mentions any other form of behaviorism.

In contrast, Stoljar's version of the perfect actor argument targets grounding behaviorism. Given this, it *does* rely on a controversial move from an epistemic premise to an ontological conclusion. And it is just not obvious that this argument succeeds, especially when its similarity to the epistemic arguments is highlighted. If Stoljar is right that the PCS entails that this version of the perfect actor argument fails, then then that gives us a principled reason to think that the perfect actor argument does fail. Of course, if there is something independently wrong with the PCS, then it should be rejected. But it should not be rejected simply because it entails that an already dubious-looking argument fails.

To be clear, my claim is just that *Stoljar's* argument against grounding behaviorism fails, not that every argument against grounding behaviorism fails. Grounding behaviorism is an extremely implausible view, and there is good independent reason to reject it. For example, it is inconsistent with our best psychological theories. Proponents of the PCS do not need to deny this. They just need to deny that one specific argument against grounding behaviorism succeeds.

2.6.4 Chalmers' dilemma

Chalmers (2007) poses a dilemma which aims to show that no version of the PCS succeeds. I discuss Chalmers' dilemma at length in chapter 3 and argue that it fails.

2.7 Conclusion

My main aim in this chapter has been to clarify the PCS and indicate which versions of it are the most promising. I argued that the analyzability version,

along with the closely related recognitional and quotational versions, succeed if the Analysis Principle is true. The recognitional and quotational versions of the PCS may also succeed if more narrowly-tailored principles about recognitional or quotational concepts are true. The other versions of the PCS, I argued, are vulnerable to serious objections.

In this chapter I also considered some general objections to the PCS (that it entails that there are too many epistemic gaps, that it overgeneralizes) and argued that they fail. There was one objection—Chalmers' dilemma—which I mentioned but did not respond to, since it requires a longer response. Chalmers' dilemma is the topic of the next chapter.

Chapter 3

Chalmers' dilemma

3.1 Introduction

As we saw in chapter 2, proponents of the PCS make arguments of the following form:

(1) *Attribution:* Phenomenal concepts are ϕ.

(2) *Explanation:* The fact that phenomenal concepts are ϕ explains the epistemic gap.

(3) *No evidence:* The fact that phenomenal concepts are ϕ is no evidence of an ontological gap.

(4) $\therefore$ The epistemic gap is no evidence of an ontological gap.

Different versions of the PCS make different substitutions for 'ϕ'. In chapter 2 I considered five main versions of the PCS, including the analyzability version which substitutes 'non-analyzable' for 'ϕ'. One way to object to the PCS is to work through these different versions of it and argue that each one fails. A different,

more ambitious approach is to argue on general grounds that no version of the PCS can succeed. This is the approach that Chalmers (2007) takes. He raises a dilemma which aims to show that no matter what ϕ is, the PCS is guaranteed to fail.

In this chapter I explain Chalmers' dilemma and say how I think proponents of the PCS should respond to it. I will argue that the dilemma poses no serious problem for proponents of the PCS, and that proponents of the PCS do not need to make any concessions to anti-materialists in response to it. This is in contrast to how some notable proponents of the PCS have responded to Chalmers' dilemma. For example, Balog (2012b) responds to Chalmers' dilemma by conceding that there is no non-question-begging way to defend the PCS from it.[1] If I am right, then defenders of the PCS do not need to make any such concession in responding to Chalmers' dilemma.[2]

3.2 The dilemma stated

Let P be the conjunction of physical truths, and C be the claim that our phenomenal concepts have some property, ϕ. (Later C will be refined slightly, but this simplified version is good enough for now.[3]) C corresponds to premise (1) of the PCS. For example, if ϕ is the property of being non-analyzable then C is the claim that phenomenal concepts are non-analyzable. Either P&$\sim$C is conceivable, or P&$\sim$C is not conceivable. These are the two horns of Chalmers' dilemma. The dilemma, then, goes like this:

[1] She also thinks, however, that there is no non-question-begging way for anti-materialists to defend the epistemic arguments from the PCS. See §3.3 below.

[2] For other responses to Chalmers' dilemma, see Papineau (2007), Carruthers and Veillet (2007), and Diaz-Leon (2010). Below I will indicate why I am dissatisfied with these responses.

[3] See pp. 62-3 below. The official version of C will say that our *quasi*-phenomenal concepts are ϕ, where quasi-phenomenal concepts are as defined below.

> Horn 1: P&∼C is conceivable. Then P does not explain C. In other words, C is not physically explicable. And if C is not physically explicable, then C is evidence of an ontological gap. But premise (3) of the PCS says that C is *no* evidence of an ontological gap. So (3) is false.
>
> Horn 2: P&∼C is not conceivable. Then, for complicated reasons explained below, C cannot explain the epistemic gap. But premise (2) of the PCS says that C *does* explain the epistemic gap. So (2) is false. (Chalmers 2007, pp. 173-4)

The upshot is that either (2) is false or (3) is false. In either case, the PCS is unsound.

We can formalize this argument as follows:

(D1) If P&∼C is conceivable, then (3) is false.

(D2) If P&∼C is not conceivable, then (2) is false.

(D3) ∴ Either (3) is false or (2) is false.

The argument is valid, so proponents of the PCS must deny either (D1) or (D2). Denying (D1) amounts to taking horn 1, and denying (D2) amounts to taking horn 2.

3.3 Horn 1: P&∼C is conceivable

On horn 1 of the dilemma, P&∼C is conceivable. Proponents of the PCS who take this horn must deny (D1). (D1) says that if P&∼C is conceivable, then (3) is false. That is, if P&∼C is conceivable, then the fact that phenomenal concepts

are ϕ is evidence of an ontological gap. The above argument for (D1) relied on two claims. First, that if P&$\sim$C is conceivable, then C is not physically explicable. Second, that if C is not physically explicable, then C is evidence of an ontological gap. These two claims straightforwardly entail (D1), so those who deny (D1) must deny one of these claims.

Denying the first claim is not an option. Chalmers has a very demanding conception of explanation, according to which A explains B only if A&$\sim$B is not conceivable. He explains that on this conception of explanation,

> explanation makes transparent why some high-level truth obtains, given that certain low-level truths obtain. If it is conceivable that the low-level truths obtain without the high-level truths obtaining, then this sort of transparent explanation will fail. (2007, p. 174)

Of course, there are less demanding conceptions of explanation. (For example, on a different conception of explanation the spark explains the fire even though it is conceivable that there is a spark and no fire.) But this is one coherent conception of explanation, and it is the one that Chalmers is working with. Given this conception of explanation, it is trivially true that if P&$\sim$C is conceivable, then C is not physically explicable.

So if a proponent of the PCS wants to deny (D1), then they must deny the second claim—the claim that if C is not physically explicable, then C is evidence of an ontological gap. This is to deny that that a certain *epistemic gap* between P and C is evidence of an ontological gap. Given such a demanding conception of explanation, this response might seem promising. After all, if so much is required for explanation, then maybe a failure of physical explanation is no evidence of an ontological gap. Nonetheless, in the context of the debate over the epistemic arguments, this response is arguably question begging.

Questions about the relation between epistemic and ontological gaps are precisely what is at issue in debates over the epistemic arguments. The PCS is supposed to show that a certain epistemic gap (the epistemic gap between physical and phenomenal truths) is no evidence of an ontological gap. Given this, it would arguably be question begging for proponents of the PCS to simply assume that the epistemic gap between P and C is no evidence of an ontological gap.[4] After all, anti-materialists think that epistemic gaps *are*, in general, evidence of ontological gaps. The PCS is supposed to show that they are wrong.

Perhaps proponents of the PCS do not need to *assume* that the epistemic gap between P and C is no evidence of an ontological gap. They can try to run a version of the PCS to *argue* for this claim.[5] After all, this is precisely the sort of claim that the PCS is designed to support. But remember that Chalmers' dilemma is an argument that the PCS fails. So in defending the PCS from Chalmers' dilemma, it is arguably question begging to simply assume that the PCS succeeds. It is not news that *if the PCS succeeds*, then Chalmers' dilemma fails. Given all this, I think that taking horn 1 of the dilemma is not a promising option for proponents of the PCS.

One point of clarification before moving on. As Chalmers notes (2007, p. 175), it might seem like proponents of the PCS must take this horn of the dilemma. This is because C is a claim about *phenomenal* concepts. A concept is a phenomenal concept only if it refers to a phenomenal state. So C is trivially false if there are no phenomenal states, since then there are no phenomenal concepts to have the

[4]This is a point that Balog (2012b) stresses in her defense of the PCS. Despite this, Balog thinks that the best option for proponents of the PCS is to deny (D1) and beg the question against anti-materialists. (She argues that anti-materialists must also beg the question against proponents of the PCS, and so it is not so bad if proponents of the PCS beg the question against them.)

[5]They could do this by arguing that our concepts of phenomenal concepts, like phenomenal concepts themselves, have the property ϕ. Alternatively, they could argue that our concepts of phenomenal concepts have some *different* property, ψ, that also explains epistemic gaps and is no evidence of ontological gaps.

property ϕ that C attributes to them. So if it is conceivable that P holds but there are no phenomenal states (as proponents of the PCS admit), then it follows that P&$\sim$C is also conceivable—as on horn 1 of the dilemma.

To avoid this result,[6] Chalmers stipulates that C be put in "topic-neutral" terms: terms that do not explicitly mention phenomenal concepts. He introduces the notion of a *quasi-phenomenal* concept—roughly, a concept that is just like a phenomenal concept, except that it may or may not refer to a phenomenal state. C will then say not that our phenomenal concepts are ϕ, but rather that our quasi-phenomenal concepts are ϕ. Every phenomenal concept is also a quasi-phenomenal concept, but there may be quasi-phenomenal concepts that are not also phenomenal concepts. So zombies might have quasi-phenomenal concepts that are not also phenomenal concepts. Given this, proponents of the PCS are not forced to say that P&$\sim$C is conceivable.

3.4 Horn 2: P&$\sim$C is not conceivable

On horn 2 of the dilemma, P&$\sim$C is not conceivable. Proponents of the PCS who take this horn must deny (D2). (D2) says that if P&$\sim$C is not conceivable, then (2) is false. That is, if P&$\sim$C is not conceivable, then the fact that phenomenal concepts are ϕ does not explain the epistemic gap. Chalmers' argument for (D2) is complicated, so it will help to start with two preliminary notes.

First, recall Chalmers' demanding conception of explanation from the discussion of horn 1: if A&$\sim$B is even conceivable, then A does not explain B. Given

[6]Why does Chalmers want to avoid this result? I take that it his idea is this. If it turns out that C is physically inexplicable only because phenomenal states are physically inexplicable, then this is not a troubling result for proponents of the PCS. Proponents of the PCS are happy to admit that phenomenal states are physically inexplicable. So long as there is nothing *additionally* inexplicable about phenomenal concepts, then the concepts themselves should not be taken as evidence of an ontological gap.

this conception of explanation, if it is conceivable that C obtains but there is nothing that corresponds to the epistemic gap between physical and phenomenal truths, then C does not explain this epistemic gap. So Chalmers' basic strategy in defending (D2) is to argue that if P&∼C is not conceivable, then it *is* conceivable that C obtains but there is no such epistemic gap.

Second, Chalmers' argument for (D2) rests on the assumption that a zombie world is conceivable. A zombie world is one that is physically identical to the actual world in which no one is conscious. This assumption will be granted to him by proponents of the PCS, since their strategy is to concede that such epistemic gaps (including the conceivability gap) exist but to argue that they are no evidence against materialism. So this assumption does not beg the question against proponents of the PCS.

Chalmers' argument for (D2) then goes like this. Suppose that P&∼C is not conceivable. Then P a priori entails C (that is, the material conditional 'if P, then C' is a priori).[7] Then the conceivable zombie world in which P obtains is also one in which C obtains. But, Chalmers claims, it is conceivable that these zombies are not faced with any epistemic gap that corresponds to the epistemic gap between physical and phenomenal truths that we are faced with. So it is conceivable that C obtains but there is no such epistemic gap. So, given Chalmers' demanding conception of explanation, C does not explain the epistemic gap that we face. Discharging the starting assumption, we get (D2).

It might help to work through things less abstractly. Suppose that C says that our phenomenal concepts are non-analyzable, as on the version of the PCS that I conditionally defended in chapter 2. If it is conceivable that zombies also

[7]This follows because a sentence is conceivable in the relevant sense iff its negation is not a priori. So if a sentence, such as P&∼C, is *not* conceivable, then its negation *is* a priori. Note also that ∼(P&∼C) is logically equivalent to the material conditional 'if P, then C'. (I am being sloppy with use and mention in this chapter.)

have non-analyzable concepts but are not faced with any epistemic gap (for example, they *can* infer propositions involving those non-analyzable concepts from the physical truths), then the mere fact that *our* phenomenal concepts are non-analyzable cannot explain the epistemic gap that we face. The argument in the previous paragraph is just a more technical and general way of spelling out this point.

That, then, is Chalmers' argument for (D2). The key claim is that it is conceivable that zombies satisfy C but are not faced with any epistemic gap that corresponds to the epistemic gap between physical and phenomenal truths that we face.[8] I think that Chalmers' argument for (D2) fails because this claim is false: it is *not* conceivable that zombies satisfy C but fail to face such an epistemic gap (at least, provided that zombies satisfy certain background conditions).[9] This is what I try to show in the next, rather long section.

For ease of presentation, in what follows I will stop qualifying talk of zombies with a conceivability operator. So I will say things like "zombies are X" when I mean "it is *conceivable* that zombies are X". Chalmers (2007) also adopts this shorthand. I will also assume, for the sake of argument, that zombies satisfy C—as they do on this horn of the dilemma. Sometimes for clarity I will make this assumption explicit, but often I will leave it implicit. So I will say things like "zombies face a corresponding epistemic gap" when I mean "*if zombies satisfy C*, then they face a corresponding epistemic gap".

Note on whether zombies can satisfy C: It might then be wondered whether zombies can (conceivably) satisfy C in the first place. This does not challenge

[8]When I say that zombies satisfy C, I mean that their schmenomenal concepts are ϕ. I am following Chalmers (2007) in adopting this shorthand.

[9]Carruthers and Veillet (2007) and Diaz-Leon (2010) also respond to Chalmers' dilemma in this way. But their defense of this key claim (that it is not conceivable that zombies fail to face an epistemic gap) is extremely brief—Carruthers and Veillet devote only a single paragraph to its defense (pp. 222-3) and Diaz-Leon devotes only a single page to its defense (p. 945). I will defend it more thoroughly.

my claim that *if* zombies satisfy C, *then* they face a corresponding epistemic gap. But it does call into question whether it is even open to proponents of the PCS to take this horn of Chalmers' dilemma. If zombies cannot satisfy C, then P&$\sim$C is conceivable, and proponents of the PCS are stuck taking the other horn of Chalmers' dilemma—which, I suggested, has significant drawbacks. So before moving on I will say a bit about this concern.

There are certainly those who would deny that zombies can satisfy C. For example, McGinn (1989) and Searle (1992) believe that consciousness is necessary for intentionality, including the grasping of concepts and the holding of beliefs. Searle calls this 'the Connection Principle'. If this principle is a priori—if it is a priori something can grasp concepts or have beliefs only if it is conscious—then it is inconceivable that zombies satisfy C or have beliefs.

An assessment of the Connection Principle is outside the scope of this chapter.[10] Here I will only note that the Connection Principle is currently unpopular among philosophers of mind. Explaining things like our concepts and beliefs (modulo their phenomenology) is generally seen as an "easy problem", in Chalmers' (1995) sense, in contrast to the "hard problem" of explaining consciousness. So I think that most philosophers of mind (including Chalmers himself) will grant that this horn of Chalmers' dilemma is at least open to take, whether or not taking it vindicates the PCS.

[10]See Kriegel (2003) for a more recent discussion of it.

3.5 The epistemic situation of zombies

3.5.1 The correspondence thesis stated

Again, my claim is that zombies *do* face an epistemic gap that corresponds to the epistemic gap between physical and phenomenal truths that we face, provided that they satisfy certain background conditions. Call this 'the correspondence thesis'. I need to say a bit to explain the correspondence thesis before defending it.

First, the background conditions are the conditions that you must satisfy to face the relevant epistemic gaps. Most notably, you must be able to form beliefs. And to face a positive conceivability gap, you must be able to imagine things. If zombies fail to satisfy these background conditions, then the correspondence thesis is vacuously true. (So if zombies cannot imagine anything, the correspondence thesis is not refuted by the fact that zombies fail to face a corresponding positive conceivability gap.) Below I will explain why, even when qualified in this way, the correspondence thesis is strong enough to rebut horn 2 of Chalmers' dilemma.

Second, the correspondence thesis is not an identity thesis: it says only that zombies face an epistemic gap that *corresponds* to the epistemic gap that we face, not that they face *exactly the same* epistemic gap as the one we face. An identity thesis would arguably be false. This is because it is plausible that the beliefs of our zombie counterparts differ from our beliefs. In particular, it is plausible that they do not have phenomenal beliefs—beliefs about phenomenal states.[11] And if they do not not have phenomenal beliefs, then they are not faced

[11]This is true on various accounts of phenomenal concepts. For example, take the concept <pain>. On the recognitional account of phenomenal concepts, <pain> refers to whatever typically triggers its applications. Since there is no pain in the zombie world, it cannot trigger the applications of any concept. So no recognitional concept in the zombie world can refer to pain.

with an epistemic gap between physical truths and phenomenal beliefs. Rather, at best they will be faced with an epistemic gap between physical truths and *schmenomenal* beliefs.

In general, two epistemic gaps, G_1 and G_2, correspond to each other iff the beliefs that make-up G_1 correspond to the beliefs that make-up G_2. I understand correspondence between beliefs roughly as Chalmers (2007) does. The basic idea is that corresponding beliefs play the same role in the agents' cognition, are expressed by the same utterances, and play the same role in explaining and predicting the agents' behavior. Corresponding beliefs are not necessarily identical because they might differ in content.

One way to get a sense of this correspondence between beliefs is to think about Twin Earth: a planet identical to the Earth apart from the fact that the clear, tasteless liquid on it is XYZ and not H_2O. When I say 'a glass of water is on the table' and my Twin Earth counterpart says the same thing, we express different beliefs: I express a belief about H_2O, and he expresses a belief about XYZ. But despite differing in content, these beliefs correspond to each other in the relevant sense: they play the same role in our cognition, we express them using the same utterances, and they play the same role in explaining and predicting our behavior.

Suppose, then, that zombies satisfy the relevant background conditions and have schmenomenal beliefs. Then the correspondence thesis is true iff zombies face an epistemic gap between the physical truths and their schmenomenal beliefs. For example, let Z be the proposition that my zombie counterpart asserts when he says 'I am in pain'. If the correspondence thesis is true, then there is an epistemic gap in the zombie world between the physical truths and Z: knowing

On the quotational account of phenomenal concepts, <pain> refers to whatever mental state it is constituted by. Since there is no pain in the zombie world, it cannot constitute any concept. So no quotational concept in the zombie world can refer to pain.

all the physical truths does not put my zombie counterpart in a position to know Z, and it is (negatively[12]) conceivable to him that the physical truths hold but Z fails to hold.

It is worth highlighting what the correspondence thesis does *not* claim. It does not claim that Z, or any other schmenomenal proposition believed by zombies, is true. It does not claim that my zombie counterpart's belief that Z is justified or constitutes knowledge. It does not claim that the entire "epistemic situation" (as Chalmers 2007 puts it) of my zombie counterpart with respect to consciousness is identical to my own. Chalmers thinks that such a limited correspondence thesis is too weak to save the PCS from horn 2 of his dilemma. But I will argue in the next subsection that the correspondence thesis, as characterized, is strong enough to save the PCS. After that, I will finally get to the defense of the thesis.

3.5.2 What must C explain?

For the purposes of the PCS, C only has to explain the epistemic gap between physical and phenomenal truths. So it only has to explain why knowing the physical truths does not put you in a position to know phenomenal truths, and why it is conceivable that the physical truths hold without phenomenal truths holding. If C can explain these things, then—barring any further problems with the PCS—the PCS succeeds in undermining the epistemic arguments.[13]

Given this, to defend the claim that C explains the epistemic gap, it is enough to defend the limited correspondence thesis that I defined above. If zombies satisfy C but have beliefs with different content, different truth values, or different

[12] As I noted above, zombies arguably lack the power of imagination and so cannot *positively* conceive of anything.

[13] Diaz-Leon (2010, p. 943), in her defense of the PCS from Chalmers' dilemma, also claims that this is all C has to explain. Carruthers and Veillet (2007, p. 222) and Papineau (2007, pp. 140-3) seem to disagree. In their defenses of the PCS from Chalmers' dilemma, they argue that zombies' schmenomenal beliefs are true.

epistemic status than our phenomenal beliefs, then Chalmers' argument might show that C cannot explain any of these features of our beliefs. But so long as zombies face an epistemic gap between the physical truths and the propositions expressed by their schmenomenal beliefs, C's explanation of the epistemic gap that we face is not undermined.

Similarly, if zombies satisfy C but fail to face an epistemic gap simply because they fail to satisfy the background conditions necessary for facing an epistemic gap in the first place (say, because they cannot form beliefs or imagine things), then this might show that C cannot explain why we satisfy these background conditions. But this does not undermine C's explanation of why those like us who *do* satisfy the relevant background conditions face an epistemic gap.

To sum up: for the purposes of the PCS, C does not have to explain (i) the *content* of our phenomenal beliefs, (ii) the *truth* of our phenomenal beliefs, (iii) our phenomenal *knowledge*, or (iv) the *background conditions* necessary for us to face an epistemic gap. As I noted above, Chalmers (2007) seems to disagree. He seems to think that, for the purposes of the PCS, C *does* have to explain (ii) and (iii). (He tacitly concedes that C does not have to explain (i) and is silent about (iv).) This is the crux of his disagreement with those who respond to his dilemma as I do. So in the rest of this subsection, I will say more to defend my claims about what C does not have to explain. I will consider each point in turn.

The content of phenomenal beliefs

I noted above that the beliefs of our zombies counterparts likely differ from our beliefs. In particular, zombies likely lack phenomenal beliefs. At best, they will have "schmenomenal" beliefs with content that differs from our phenomenal beliefs. So there is at least one feature of the epistemic gap we face that C alone

cannot explain: that it consists in part in phenomenal beliefs as opposed to, say, schmenomenal beliefs.[14] But this does not undermine C's explanation of why, *given that we have the beliefs that we do have*, we face an epistemic gap that involves those beliefs.

Here is an analogy that might help explain this point. Suppose, as Block and Stalnaker (1999) argue, that there is an epistemic gap between the physical truths and truths that involve the concept <water> (for example, 'there is water on Earth').[15] And suppose that there is this epistemic gap because <water> is non-analyzable. Now consider our counterparts on Twin Earth. Their concept <twater> corresponds to our concept <water>, and is also non-analyzable. But they will not face exactly the same epistemic gap that we face, since they have no beliefs that involve the concept <water>. Instead, they will face merely a *corresponding* epistemic gap—one between the physical truths and truths that involve <twater>. Despite this, the fact that <water> is non-analyzable might still be a good explanation of the epistemic gap that we face.

The truth of phenomenal beliefs

It is debatable whether zombies' schmenomenal beliefs are true. In an appendix I argue that they are true, but this claim is controversial and its defense is unnecessary for a defense of the PCS.[16] In this chapter, I am willing to grant for

[14] C might be a good *partial* explanation of why our phenomenal beliefs have the content they do. For example, part of the explanation of the content of our phenomenal beliefs will say that phenomenal concepts are quotational, or recognitional, or whatever. But for this to be a full explanation in Chalmers' demanding sense of explanation, we would have to add that these concepts refer to phenomenal sates. Otherwise it would be conceivable that our beliefs lack the content that they in fact have (as they do in the zombie world), even if we satisfy C.

[15] Remember that by 'physical truths' I mean '*narrowly* physical truths', where these are the truths that can be stated in the language of fundamental physics (see chapter 1). 'There is water on Earth' does not count as a narrowly physical truth because, among other reasons, 'water' is not a term in the language of fundamental physics.

[16] As I noted above, Carruthers and Veillet (2007, p. 222) and Papineau (2007, pp. 140-3) seem to think that this *is* necessary for a defense of the PCS.

the sake of argument that zombies' schmenomenal beliefs are false. If zombies satisfy C but their schmenomenal beliefs are false, then C cannot explain the truth of our phenomenal beliefs. But, importantly, this does not undermine C's explanation of the epistemic gap.

Chalmers seems to think that it does, and this is the heart of his attack on my response to his dilemma. (What he says about this is open to interpretation, which is why I say only that Chalmers *seems* to think this.) Here is what he says when considering the response to his dilemma that I am defending:

> If one characterized these [epistemic] gaps in a way that were neutral on the truth of phenomenal beliefs, the [epistemic] arguments would not get off the ground. So truth value is essential to the relevant epistemic gaps. If so, then to undercut the inference from these gaps to an ontological gap, the phenomenal concept strategy needs to show how the relevant truth-involving epistemic gaps are consistent with physicalism. The earlier strategies do not do this, so they do nothing to undercut the inference from the epistemic gap to an ontological gap. (2007, p. 183)

One of the "earlier strategies" claimed that C needs to explain only the inferential disconnection between the physical truths and our phenomenal beliefs.

I actually agree with most of this passage. Chalmers is right that a good explanation of the epistemic gap that we face must explain why knowing the physical truths does not put you in a position to know phenomenal *truths*. The explanation of why there is an epistemic gap between physical truths and phenomenal beliefs would be trivial if no phenomenal beliefs were true. After all, truths do not entail falsehoods, and nothing puts you in a position to know a falsehood. There is something interesting to explain about this epistemic gap

only because there are *truths* that are not entailed by the physical truths.

So I agree with Chalmers that the following is a constraint on a good explanation of the epistemic gap between the physical truths and our phenomenal beliefs:

> *The Constraint (truth version):* The explanation of the epistemic gap between the physical truths and our phenomenal beliefs must explain how there could be an inferential disconnection between them even if our phenomenal beliefs are true.

But towards the end of the above passage, Chalmers seems to assume that C satisfies the Constraint only if C explains the truth of our phenomenal beliefs. This is why he says that the "earlier strategies", which do not explain the truth of our phenomenal beliefs, do not satisfy the Constraint. And in the next paragraph he goes on to consider materialist attempts to explain the truth of our phenomenal beliefs, and argues that none can succeed. ("Perhaps proponents could augment their explanation of the narrow epistemic situation with an additional element that explains why the relevant beliefs are true and qualify as knowledge…")[17]

This, I think, is a big mistake. C can satisfy the Constraint even if it does not explain the truth of our phenomenal beliefs. Here is an analogy that might help explain this point. Suppose that a given poison can kill even a previously healthy person. Then my receiving the poison explains why, even though I was previously healthy, I died. But my receiving the poison does not explain *why I was previously healthy*. And the explanation of my death is not undermined by the fact that there is a conceivable world in which *unhealthy* people receive the poison. In that world, it just happens that the deaths of the relevant people may be over-determined.

[17] Diaz-Leon (2010, pp. 954-7) also quotes the above passage and reads Chalmers as claiming that, for the purposes of the PCS, C must explain the truth of our phenomenal beliefs.

Similarly, that we satisfy C explains why, even though our phenomenal beliefs are true, they cannot be inferred from the physical truths. But it does not explain *why our phenomenal beliefs are true*. And the explanation of the epistemic gap is not undermined by the fact that there is a conceivable world in which zombies satisfy C but have beliefs corresponding to our phenomenal beliefs that are *false*. In that world, it just happens that the epistemic gap faced by zombies is over-determined (since, in general, falsehoods do not follow from truths).

Phenomenal knowledge

Since I am willing to grant for the sake of argument that zombies' schmenomenal beliefs are false, I am also willing to grant that their beliefs do not constitute knowledge: if they are not true, then a fortiori they do not constitute knowledge. And if zombies satisfy C but lack knowledge of the relevant propositions, then C cannot explain our knowledge of the corresponding phenomenal propositions. But as before, this is not a problem for the PCS, because it does not undermine C's explanation of the epistemic gap.

Chalmers seems to disagree:

> Whereas the inferential disconnection strategy may physically explain an inferential disconnection between physical and phenomenal *beliefs*, the antiphysicalist's crucial epistemic gap involves a disconnection between physical and phenomenal *knowledge*. This strategy does not help to reconcile this crucial epistemic gap with physicalism, so it lends no support to type-B materialism. At best, it shows that zombie-style analogs of phenomenal beliefs (inferentially disconnected from physical beliefs) are compatible with physicalism, but this is something that we knew already. (2007, pp. 183-4)

As before, I agree with much of this passage. Since we in fact have phenomenal knowledge, a good explanation of the epistemic gap must explain why, even if we have phenomenal knowledge, knowing the physical truths does not put us in a position to know phenomenal truths. If the explanation failed to apply to those with phenomenal knowledge, then it would fail to apply to us and the epistemic gap that we face, and so would be useless in explaining this gap. So I agree with Chalmers that the following is a constraint on a good explanation of the epistemic gap between the physical truths and our phenomenal beliefs:

> *The Constraint (knowledge version):* The explanation of the epistemic gap between the physical truths and our phenomenal beliefs must explain how there could be an inferential disconnection between them even if our phenomenal beliefs constitute knowledge.

But in the above passage, Chalmers seems to assume that C satisfies the Constant only if C explains our phenomenal knowledge. And, as before, I think this is a big mistake. The idea is the same as before. That I was poisoned can explain (i) why I died even though I was previously healthy, without explaining (ii) why I was previously healthy. Similarly, that we satisfy C can explain (i) why we face an epistemic gap even though we have phenomenal knowledge, without explaining (ii) why we have phenomenal knowledge. So that zombies satisfy C but lack schmenomenal knowledge might show that C cannot explain our phenomenal knowledge, but it does not show that C cannot explain the epistemic gap.

Background conditions

It is debatable whether zombies satisfy the background conditions for facing all of the epistemic gaps that we face. If zombies satisfy C (as they do on this horn

of the dilemma), then they will probably at least have beliefs—one background condition for facing any epistemic gap.[18] But there are other background conditions that they might fail to satisfy. Most notably, they arguably lack the power of imagination. This is a background condition for facing a *positive* conceivability gap: if you cannot imagine anything, then a fortiori you cannot imagine a world in which one set of truths holds but another set fails to hold.

But this does not undermine C's explanation of the positive conceivability gap that we face. Here is an analogy that might help explain this point. That someone is weak-willed explains why they waste time on Facebook. It does not explain why they have internet access—a background condition for wasting time on Facebook. Nonetheless, this explanation of their Facebook use is a good one. Similarly, C might explain why we face a positive conceivability gap even if it does not explain why we can imagine things—a background condition for facing a positive conceivability gap.

So all proponents of the PCS need to show is that *if* zombies satisfy the relevant background conditions, then they face a corresponding epistemic gap. If zombies fail to satisfy the background conditions, then we would not expect them to face a corresponding epistemic gap *even if* C explains the epistemic gap we face. Similarly, if someone lacks internet access, then we would not expect them to waste time on Facebook *even if* being weak-willed explains why those who have internet access waste time on Facebook.

[18] Recall that C is a claim about our quasi-phenomenal concepts: that they have a particular property, ϕ. So if zombies satisfy C, then they at least have quasi-phenomenal *concepts*. Given this, it is hard to see how they could also lack corresponding *beliefs* that involve those concepts.

3.5.3 The correspondence thesis defended

Again, the correspondence thesis says that zombies face an epistemic gap that corresponds to the epistemic gap between physical and phenomenal truths that we face, provided that they satisfy certain background conditions. In particular, it claims that they face (i) a *knowledge gap*, so knowing the physical truths does not put zombies in a position to know the propositions expressed by their schmenomenal beliefs, and (ii) a *conceivability gap*, so it is (negatively) conceivable to them that the physical truths hold but the propositions expressed by their schmenomenal beliefs are false.

Given how weak the correspondence thesis is, it should be relatively uncontroversial. Even Chalmers seems to concede it: he says that "it is plausible that a zombie's physical and quasi-phenomenal beliefs are no more inferentially connected than a conscious being's beliefs" (2007, p. 183). But it is still worth going through the argument for it. I will assume for the sake of argument that zombies satisfy the background conditions necessary for facing epistemic gaps. If they do not satisfy these background conditions, then the correspondence thesis is vacuously true.

To start, note that zombies are functional duplicates of us.[19] This requires, at the very least, that their concepts play the same conceptual roles as our corresponding concepts. Part of the conceptual role played by a concept is the valid inferences it figures in: if two concepts play the same conceptual role, then the valid inferences involving the one concept correspond to the valid inferences involving the other concept. For example, if the concepts expressed by F and G play the same conceptual role, then P entails Fx iff P entails Gx.

[19] *Pace* Shoemaker (1975), who argues that it is inconceivable that a creature that shares our functional states lacks consciousness. As a concession to anti-materialists, I will assume that Shoemaker is wrong about this. If Shoemaker is right, then premise (E) of the epistemic arguments is false, and so the epistemic arguments fail for reasons independent of the PCS.

Now I am assuming that there is an epistemic gap between physical and phenomenal truths. (This is a concession to anti-materialists. If there is no such epistemic gap, then the epistemic arguments fail for reasons independent of the PCS.) So if there is no corresponding epistemic gap between physical truths and schmenomenal propositions, then phenomenal and schmenomenal concepts do not play the same conceptual role. Since phenomenal and schmenomenal concepts *do* play the same conceptual role, there is an epistemic gap between physical and schmenomenal truths.[20]

As I noted above, since zombies arguably lack the power of imagination, they will arguably be unable to imagine anything—a background condition for facing a positive conceivability gap. But this does not undermine the correspondence thesis. It *would* undermine the correspondence thesis if zombies could imagine things, but just happen to be able to see a contradiction in any sufficiently detailed imagined situation in which P&$\sim$Z (where Z expresses a schmenomenal proposition). But it does not undermine the correspondence thesis if zombies fail to face a positive conceivability gap simply because they cannot imagine anything at all, and so a fortiori cannot imagine a coherent situation in which P&$\sim$Z.

That, then, is my defense of the correspondence thesis. Recall that the correspondence thesis defends the PCS from horn 2 of Chalmers' dilemma. Chalmers denied the correspondence thesis in arguing for (D2). He argued that zombies do not face any epistemic gap that corresponds to the epistemic gap that we face, and so if C is true in the zombie world then it cannot explain the epistemic gap that we face. If the correspondence thesis is true then this argument is unsound, since zombies *do* face an epistemic gap that corresponds to the epistemic gap that we face (provided that they satisfy the relevant background conditions).

[20] Balog (1999, pp. 514-5) and Diaz-Leon (2010, p. 945) also argue along these lines.

3.6 Conclusion

I have argued that Chalmers' dilemma fails: it gives us no reason to reject, or even significantly weaken, the PCS. Of course, this alone does not mean that the PCS succeeds. It just means that if it fails, then it does not fail on account of Chalmers' dilemma. But in conjunction with the conditional defense of the PCS from chapter 2, I think that this puts the PCS on solid ground.

There are still some lose ends to deal with. I argued in chapter 2 merely that *if* the Analysis Principle is true, then the PCS succeeds. So a natural question is what follows if the Analysis Principle is false. I address this question in chapter 4. I try to show that anti-materialists need to rely on the Analysis Principle just as much as proponents of the PCS do, and so if it is false then the epistemic arguments fail for reasons independent of the PCS. Another natural question is whether the Analysis Principle is true. I assess the Analysis Principle in chapter 5.

Chapter 4

A tension in the epistemic arguments

4.1 Introduction

In chapter 2 I presented a conditional defense of the PCS—conditional, because I argued that *if* the Analysis Principle is true, then the PCS succeeds. But I suggested that materialists do not need to take a stand on the truth or falsity of the Analysis Principle. This is because, I claimed, anti-materialists themselves must appeal to the Analysis Principle in defending the epistemic arguments. If this is right, then either the Analysis Principle is true and the PCS succeeds, or the Analysis Principle is false and the epistemic arguments fail for reasons independent of the PCS. Either way, the upshot is good for materialists.

In this chapter I develop this argument in more detail. A little more precisely, what I will argue is this. To successfully defend (E), the claim that there is an epistemic gap, anti-materialists must appeal to considerations about phenomenal

concepts and use the Analysis Principle to infer from these considerations that there is an epistemic gap. This, in fact, is how many anti-materialists *do* argue for (E). But they make other arguments for (E), too. What I will try to show is that the *only* argument for (E) that stands a chance of succeeding (apart, perhaps, from an independent argument against materialism[1]) is the one that relies on the Analysis Principle.

At the same time, anti-materialists can defend (E→O)—the claim that the epistemic gap entails an ontological gap—from two major objections only by *denying* the Analysis Principle. One objection is the PCS. As I argued in chapter 2, if the Analysis Principle is true, then the analyzability version of the PCS succeeds, and so we have no reason to accept (E→O). Another objection (suggested by Block and Stalnaker 1999) appeals to the Analysis Principle in arguing that epistemic gaps are rampant, and so are no evidence of ontological gaps.

So anti-materialists might be might be able to appeal to the Analysis Principle and defend (E). And they might be able to deny the Analysis Principle and defend (E→O). But they cannot do both at the same time, since these defenses are inconsistent. This does not mean that (E) and (E→O) *themselves* are inconsistent. Maybe they are both be true, even though there is no good way to defend both of them. So it does not follow from this that the epistemic arguments are unsound. Still, if there is no good defense of both (E) and (E→O), then the epistemic arguments should not seriously worry materialists.

[1] See §4.2.4 below.

4.2 Arguments for (E)

4.2.1 Appeals to intuitions

(E), again, is the claim that there is an epistemic gap between physical and phenomenal truths. It is worth emphasizing that (E) is committed to the existence of an epistemic gap that would exist even if we were ideal reasoners who knew all the truths of a completed physics. A *shallow* epistemic gap, one that existed merely because we are non-ideal reasoners or because we are ignorant of some physical truth, poses no threat to materialism. It is only a *deep* epistemic gap, one that does not arise from our cognitive limitations or physical ignorance, that poses a threat to materialism.[2]

The first sort of argument for (E) that I will consider appeals to intuitions. One version of this argument goes like this. We have the intuition that Mary would learn something when she leaves the black-and-white room, so probably she *would* learn something when she leaves the black-and-white room. Another version goes like this. We have the intuition that zombies are conceivable, so probably zombies *are* conceivable.[3] I think that most of us do have the intuitions that these arguments claim that we have. But I deny that these intuitions give us good reason to accept (E). This is for two reasons.

First, intuitions are not always a good guide to what is knowable to or

[2]By default, I will mean 'deep epistemic gap' if I simply say 'epistemic gap' without qualification. But in what follows, I will often be explicit about which sort of epistemic gap I am talking about.

[3]Anti-materialists rarely argue for (E) in this way. For example, Jackson (1986) simply claims without argument that Mary would learn something when she leaves the black-and-white room. He does not claim that it is *intuitive* that she would learn something, and then infer from this that she would in fact learn something learn something. (Deutsch 2015 argues that philosophical arguments in general rarely rely on the claim that something is intuitive as a premise, and he considers the knowledge argument as one test case.) Still, (E) is far from obvious, and it is reasonable to challenge anti-materialists to defend it. The appeal to intuitions is one possible defense. If, however, you are already convinced that this defense of (E) fails, then you can skip to the next section, since you already accept what I trying to show in this section.

conceivable for *ideal* reasoners. It might be that one truth is derivable from another truth, but only after a huge amount of reasoning. Or it might be that there is a contradiction in an apparently conceivable scenario, but that this contradiction is revealed only after a huge amount of reasoning. Since we are non-ideal reasoners, our intuitions might fail to be sensitive to these inferential connections or contradictions. This objection to the conceivability argument is sometimes put by saying that although zombies are prima facie conceivable, they might not be ideally conceivable.[4]

The most obvious examples of prima facie conceivability without ideal conceivability come from math. For example, it is prima facie conceivable that Fermat's Last Theorem is false, but this is not ideally conceivable (since the theorem is in fact true).[5] But there are other, more challenging examples. For example, it is prima facie but not ideally conceivable that there is a complete physical duplicate of my laptop that does not have a copy of this dissertation on it: *I*, with my limited intellect, cannot see a contradiction in in that scenario, but someone smart enough would be able to see a contradiction in it.

This is not just a theoretical worry. There are historical examples of philosophers—even philosophers of mind, specifically—being mislead by their intuitions about conceivability. Stoljar (2006) gives the example of Descartes, who had the intuition that it is not conceivable that a machine have the ability to use language.[6] Descartes was a careful and reflective philosopher. But he was not an ideal reasoner, and in this case his intuitions mislead him. On more careful reflection (in light of, say, computer science) we can see that it *is* conceivable that

[4]See Chalmers (2003a, p. 119). Ideal conceivability is the same as what until now I have simply called 'conceivability'.

[5]Chalmers (2002, pp. 147, 160) discusses this sort of case and argues that it poses little problem for those who want to use conceivability as a guide to possibility.

[6]Stoljar (2006, pp. 124-5), Descartes (1637/1998, §V).

a machine have the ability to use language.[7]

This is not to say that we should, for this reason, *always* doubt our intuitions of conceivability. I have the intuition that it is conceivable that pigs fly, and this gives me good reason to think that it *is* ideally conceivable that pigs fly. That scenario is simple enough that there is little room for my intuitions to mislead me about its conceivability, even though my intellect is limited. But if a scenario is more complicated, and especially if it involves unobvious connections between lower-level and higher-level domains (as in the dissertation-laptop and mind-body cases), then our intuitions are liable to mislead us.[8]

I will respond to two objections before moving on.

Objection 1: It's not that complicated:[9] The fact that we are non-ideal reasoners gives us reason to doubt only our intuitions about extremely complicated scenarios, but the zombie scenario is not that complicated. It is true that if, to conceive of a zombie, you had to keep in mind every detail of its physical make-up (including every atom), then that *would* be extremely complicated. But you do not need to keep all these details in mind to conceive of a zombie. If consciousness is grounded in the physical, it is grounded in higher-order features of physical things. So it is just these higher-order features that you need to keep in mind when conceiving of a zombie. And doing that is not too complicated.

Reply: First, it is not obvious that only the higher-order features of some-

[7] As Stoljar notes, Descartes' intuitions about language use are not perfectly analogous to our intuitions about zombies. Descartes' intuition about language use was that something is *not* conceivable. Our intuitions about zombies are that they *are* conceivable. But the point is just that our intuitions about conceivability (be they intuitions that something *is* conceivable or intuitions that something is *not* conceivable) can be unreliable.

[8] *Mere* complexity (understood in terms of, say, the sheer number of objects or parts of objects in a given scenario) is arguably not a reason to doubt our intuitions about the scenario. For example, I have the intuition that it is conceivable that 10^{100} books can be piled on top of each other, and this intuition is arguably reliable. But in this case, there is no hard question about whether the books in some unobvious way "add up" to something else, as in the dissertation-laptop and mind-body cases.

[9] This objection comes from Chalmers (1996, p. 98).

thing's physical make-up are relevant to its phenomenal states. Maybe a computer would not be conscious even if it had all the relevant higher-order features that my brain has. This is currently an open question. Second, even higher-order features can be extremely complicated. For example, integrated information theory (IIT) says that consciousness is grounded in certain information states.[10] But although these are higher-order features of our brains, they are still extremely complicated. Third, even figuring out what the relevant higher-order features are is itself an intellectual feat. It is not as if IIT and other proposed theories of consciousness are obvious. They take an enormous amount of careful thought to develop. (Compare Descartes' ignorance of computer science, which made him think falsely that a machine cannot conceivably use language.)

Objection 2: Positive vs. negative conceivability:[11] Distinguish between negative and positive conceivability. A sentence is *negatively* conceivable iff its negation is not a priori. A sentence is *positively* conceivable iff you can imagine a situation in which it is true, and arbitrary details of this imagined situation can be filled-in without contradiction.[12] Now it is true that our intuitions that something is negatively conceivable might be misleading. That is what the Fermat's Last Theorem example shows. But our intuitions that something is *positively* conceivable are reliable. And we have the intuition that zombies are positively conceivable. So, probably, zombies really are (positively) conceivable.

Reply: It is true that we can form some not-very-detailed mental image of zombies. But for zombies to be positively conceivable, it must be possible to fill-in arbitrary details to this imagined situation without contradiction. And it is not obvious that a zombie situation will remain coherent when we fill-in details about, say, the neurons in the zombies' brains. To return to the laptop

[10]For an overview of IIT, see Fallon (2016). See also Tononi (2007).
[11]This objection comes from Chalmers (2002, pp. 145, 160-1).
[12]See Chalmers (2002, pp. 149-54) for a fuller definition of positive conceivability.

example: I can form some not-very-detailed mental image of a laptop that is physically identical to my current laptop, but which does not have a copy of this dissertation on it. But this imagined situation would not remain coherent if I filled-in details about the exact state of the hard drive. It might be objected that fine-grained details are irrelevant to the conceivability of zombies, but I addressed this point when responding to objection 1.

That, then, is the first reason why our intuitions do not give us good reason to accept (E): we are non-ideal reasoners, and there may be an unobvious epistemic connection between physical and phenomenal truths. The second reason is that we might have these intuitions merely because we are working without knowledge of a complete physics. For example, maybe if we knew all the truths of a completed physics, we would not have the intuition that Mary would learn something when she leaves the black-and-white room. Or maybe we would not have the intuition that it is conceivable that all the truths of a completed physics can hold without the phenomenal truths holding. Given that we do not know what a completed physics would look like, these possibilities are hard to rule out.

As before, this is is not to say that we should, for this reason, *always* doubt our intuitions of conceivability. For one, not every apparently conceivable scenario is characterized in terms of the actual physical truths. But even our intuitions about scenarios that are characterized in this way are not necessarily unreliable, despite our physical ignorance. For example, I have the intuition that it is conceivable that the fundamental physical truths are what they are but there is a mile-high pile of gold, and this seems to give me good reason to think that this *is* ideally conceivable (despite my ignorance of the complete physical theory).[13] But the

[13]Thanks to an anonymous reviewer for drawing my attention to this sort of case. Of course, if my physical ignorance were extreme enough, then my intuitions about the conceivability of even this scenario would be unreliable.

mind-body case differs because there are serious theories of consciousness on which unknown physical truths are relevant to consciousness.[14] In contrast, there is no serious theory of gold on which unknown physical truths are relevant to how high gold can be piled.

So I do not think that mere appeals to intuitions offer much support for (E). This does not mean that appeals to intuitions have no force. The intuitions mentioned above arguably show that there is some sort of epistemic gap between physical and phenomenal truths. But crucially, they leave open the possibility that the epistemic gap is only a *shallow* one: one that arises from our cognitive limitations or physical ignorance. And to show that (E) is true, what needs to be shown is that there is a *deep* epistemic gap between physical and phenomenal truths: one that does *not* arise from our cognitive limitations or physical ignorance.

This also does not mean that out intuitions offer *no* support to the claim there is a deep epistemic gap. They do support it slightly, but not enough to make the epistemic arguments a serious concern for materialists. As an analogy, suppose that someone were defending a highly revisionary view about computer science by appeal to the claim that there is a deep epistemic gap between the physical truths and truths about the information stored on my laptop. Our intuitions offer *some* support for the epistemic gap claim here, but not nearly enough to make this argument a serious concern for computer scientists. I claim the same thing about our intuitions about the epistemic gap between physical and phenomenal truths.

[14]This view is most commonly defended by Russellian monists, who claim that phenomenal truths are grounded in part in the intrinsic properties of matter (which, they claim, we are currently ignorant of). See, for example, Stoljar (2001), Pereboom (2011), Alter and Nagasawa (2015), and Brüntrup and Jaskolla (2016). In more recent work, Stoljar (2006, 2015) has argued that unknown physical truths may be relevant to consciousness even if these truths have nothing to do with the intrinsic properties of matter. He calls this view 'Nagelian monism'.

4.2.2 Conceptual considerations

The second sort of argument for (E) that I will consider appeals to considerations about phenomenal concepts, and I think that this sort of argument probably does succeed if the Analysis Principle is true. In its simplest form, this argument claims that phenomenal concepts cannot be analyzed into physical or functional concepts, and then infers from this that there is an epistemic gap between physical and phenomenal truths. For example, here is what Chalmers (1996) says:

> For consciousness to be [a priori] entailed by a set of physical facts, one would need some kind of analysis of the notion of consciousness... and there is no such analysis to be had (p. 104).[15]

Chalmers talks about a priori entailment specifically, but his point seems to be about epistemic gaps more generally. The idea seems to be that, in general, conceptual analysis is necessary to close epistemic gaps. This is what the Analysis Principle says:

> **Analysis Principle:** If the concepts involved in one truth differ from and cannot be analyzed into the concepts involved in another truth, then there is an epistemic gap between the two truths.

Here the epistemic gap is understood as a deep epistemic gap (again, one that cannot be closed even on ideal reflection), as opposed to a shallow one. Conceptual analysis is understood as in chapter 2, in terms of the a priority of certain sentences.

This second sort of argument for (E) avoids both of the problems faced by appeals to intuitions. If the Analysis Principle is true, then there is a principled reason for thinking that *even on ideal reflection* Mary would not be able to learn

[15] Balog (1999, p. 506) endorses a similar argument.

what it is like to see red while in the black-and-white room, and that *even on ideal reflection* we would be able to conceive of zombies. Similarly, if the Analysis Principle is true then, so long as phenomenal concepts cannot be analyzed into the concepts of a completed physics, there is a principled reason for thinking that knowing the truths of a *complete* physics does not put you in a position to know the phenomenal truths, and that it is conceivable that the truths of a *complete* physics can hold without the phenomenal truths holding.

The drawback of this sort of argument is that it relies on the Analysis Principle, which as we will see below is extremely controversial. Still, provided that the Analysis Principle is true, I think that this makes a compelling case for (E).

4.2.3 The structure and dynamics argument

Another argument for (E) is the structure and dynamics argument, first defended by Chalmers (2003a) and later developed more fully by Alter (2015). This argument rests on a distinction between purely structural and dynamic truths (for short, 'structural truths') and non-purely-structural-and-dynamic truths (for short, 'non-structural truths'). I explain this distinction below, but to start it will help to simply get the argument on the table. It can be formalized as follows:

(1) Every physical truth is structural.

(2) No structural truth a priori entails a non-structural truth.

(3) No phenomenal truth is structural.

(4) ∴ No physical truth a priori entails a phenomenal truth.[16]

[16]For various presentations of the structure and dynamics argument, see Chalmers (2003a), Alter (2015), and Stoljar (2014, 2015). As I have presented it, the argument aims to defend (E). But others have used the same sort of argument to defend other views in the philosophy of mind, such as Russellian monism. See, for example, Pereboom (2011, 2014, 2015).

If (4) is true, then there is a deep epistemic gap between physical and phenomenal truths, and (E) is also true. I will consider two versions of the argument, each of which corresponds to a different understanding of structural truths. I will argue that one version of the argument amounts to a sophisticated version of the conceptual considerations argument from §4.2.2 (and so relies on the Analysis Principle), and the other version of the argument either fails for the same reason that the appeals to intuitions from §4.2.1 fail or relies on controversial claims about the nature of phenomenal concepts.

Conceptual version (Alter)

On Alter's (2015) understanding of structural truths, they are defined in terms of a special set of expressions, which includes mathematical, logical, and nomic expressions (expressions to do with cause and effect). Call these 'structural expressions'. A *structural truth* is a truth that contains only structural expressions or expressions that can be analyzed into structural expressions. Conceptual analysis is understood as in chapter 2, in terms of the a priority of certain sentences. A *non-structural truth* is a truth that is not structural. That is, it is a truth that contains at least one expression that is not structural and cannot be analyzed into structural expressions.

On this understanding, (3) ("No phenomenal truth is structural") says that every phenomenal truth contains at least one expression that is not structural and cannot be analyzed into structural expressions. This claim should look familiar. It is very close to the claim that phenomenal concepts cannot be analyzed into physical concepts, which played a key role in the conceptual considerations argument from §4.2.2. Importantly, this is just a claim about conceptual analysis and can be justified by linguistic intuitions. It is not a metaphysical claim about

the nature of phenomenal states themselves, which would not be as easy to justify.

The key premise is (2) ("No structural truth a priori entails a non-structural truth"). On this understanding, (2) amounts to the claim that lack of conceptual analysis of a certain sort entails lack of a priori entailment. To see this, suppose that T_1 is a structural truth and T_2 is a non-structural truth. This is to suppose that T_1 contains only structural expressions and expressions analyzable into structural expressions, and that T_2 contains at least one expression that is neither structural nor analyzable into structural expressions. (2) says that it follows from these facts about conceptual analysis that T_1 does not a priori entail T_2.

Although (2) does not say anything more general about a priori entailment, the underlying idea seems to be that, in general, conceptual analysis is necessary for a priori entailment. After all, it is hard to see why lack of conceptual analysis *into structural expressions* would block a priori entailment unless lack of conceptual analysis *in general* blocked a priori entailment. Why would conceptual analysis into this one type of expressions be special? So this version of the structure and dynamics argument, like the conceptual considerations argument, arguably relies on something like the Analysis Principle. Still, I think that if the Analysis Principle is true, then this makes a compelling case for (E).

Could (2) be defended in a way that does not rely on the Analysis Principle? I do not think so. The only other attempt I can think of is to appeal to a metaphysical claim—say, that structural truths fail to fully ground non-structural truths, and so do not a priori entail structural truths. The problem with this argument is that the structural/non-structural distinction, as characterized by Alter, is a *conceptual* distinction, and it is not obvious that structural truths fail to fully ground non-structural truths. Considerations about metaphysics are, however,

important for the second version of the structure and dynamics argument that I will consider.

Metaphysical version (Pereboom)

Pereboom (2011, 2014, 2015) understands structural truths in terms of absolutely intrinsic properties. Some intrinsic properties of an object, such as its shape, are grounded in the relations that its proper parts stand in to each other. These are merely *comparatively intrinsic* properties. In contrast, an intrinsic property of an object that is not grounded in the relations that its proper parts stand in to each other is *absolutely intrinsic*. There are no uncontroversial examples of absolutely intrinsic properties, but Pereboom gives a few examples that have been suggested by philosophers (2014, p. 55).[17] A *structural truth*, then, is a truth that is fully grounded in properties that are either extrinsic or comparatively intrinsic. A *non-structural truth* is a truth that is not structural.[18] That is, it is a truth that is either fundamental or at least partly grounded in a property that is absolutely intrinsic.

On this understanding, (2) ("No structural truth a priori entails a non-structural truth") is plausible. It amounts to the uncontroversial claim that a certain ontological gap entails a corresponding epistemic gap.[19] To see this, suppose that T_1 is a structural truth and T_2 is a non-structural truth. Since T_2 is

[17]These include perfect solidity (Locke and Newton), color (Armstrong), phenomenal properties (Leibniz, Strawson, and Adams), and protophenomenal properties (Chalmers).

[18]Nico Silins pointed out in conversation that this definition has the consequence that all fundamental truths are non-structural, which might be intuitively wrong. For example, if mathematical truths are fundamental, then they count as non-structural on this definition. You could try to avoid this consequence by saying that a non-structural truth is a truth that is both (i) non-fundamental and (ii) not structural. But this seems too restrictive, since there might be fundamental phenomenal truths that intuitively should count as non-structural. In any case, this wrinkle is irrelevant to the point that I want to make in response to Pereboom's version of the structure and dynamics argument.

[19]Note that this is not the reverse claim that an *epistemic* cap entails a corresponding *ontological* gap, which *would* be controversial.

non-structural, it is at least partly grounded in an absolutely intrinsic property. Since T_1 is structural, it does not involve any absolutely intrinsic properties. So T_2 cannot be fully grounded in T_1, and there is a ontological gap between T_1 and T_2. Given this, there is also an epistemic gap between T_1 and T_2: if T_1 leaves it unsettled whether T_2 is true, then there is no way to a priori infer T_2 from T_1.[20]

The key premise is (3). On this understanding, (3) ("No phenomenal truth is structural") says that every phenomenal truth is at least partly grounded in an absolutely intrinsic property. Importantly, this is not a claim about conceptual analysis. It is a claim about the nature of phenomenal states themselves, and cannot be justified simply by linguistic intuitions. In fact, (3) comes extremely close to entailing that materialism is false. Anti-materialism follows when (3) is combined with (1) ("Every physical truth is structural"). This does not mean that (3) is false. But it makes it doubtful that arguing for (3) is much easier than simply arguing against materialism.

I can think of two arguments for (3). The first argument for (3) claims that there is an *epistemic gap* between structural and phenomenal truths, and infers from this that there is a corresponding *ontological gap*, and consequently that phenomenal truths are not grounded in the structural truths. This argument has the same structure as the epistemic arguments. There is some intuitive support for the epistemic gap premise. The problem is that the epistemic gap premise is open to the same objection that I raised to (E) this chapter: that the epistemic gap is merely shallow, not deep, and that only a deep epistemic gap would entail an ontological gap.

It might help to review the dialectic that got us here. Anti-materialists argue

[20] I am setting aside a priori truths, which generate counterexamples to this claim. For example, '2+2=4' is not grounded in 'grass is green', but 'grass is green' trivially a priori entails '2+2=4' since '2+2=4' is a priori. So ontological gaps do not *always* entail epistemic gaps, but they usually do.

against materialism by claiming that there is a deep epistemic gap between physical and phenomenal truths. This is what (E) says. They might try to defend (E) by appealing to intuitions, but I argued that these intuitions show at best that there is a shallow epistemic gap between physical and phenomenal truths. If the structure and dynamics argument is to improve on mere appeals to intuitions, then it must give us some reason to think that there is a deep epistemic gap between physical and phenomenal truths. But if (3) is defended in this way, then the structure and dynamics argument itself ultimately rests on intuitions about an epistemic gap (one between *structural* and phenomenal truths), and it gives us no reason to think that *this* epistemic gap is deep.

There is admittedly one way in which our intuitions about an epistemic gap between structural and phenomenal truths are less open to doubt than our intuitions about an epistemic gap between *physical* and phenomenal truths: worries about our physical ignorance give us no reason to doubt our intuitions about the first putative epistemic gap, since structural truths are not defined in terms of (possibly unknown) physical truths. But the other ground for doubt remains: there may be an unobvious epistemic connection between structural and phenomenal truths that we, because of our cognitive limitations, cannot easily see. This case looks relevantly similar to ones about which our intuitions are especially liable to mislead us, such as the dissertation-laptop case from §4.2.1 above.

That, then is, one argument for (3). The other argument for (3) comes from Pereboom (2014). It goes like this. Our phenomenal concepts represent phenomenal properties as absolutely intrinsic. These representations are accurate. So phenomenal properties are absolutely intrinsic. So they are not fully grounded in purely structural properties, and (3) is true.

I do not think this argument is sound. Stoljar (2014, 2015) responds at more length to this argument. But briefly, both its premises are questionable. First, it is not obvious that our phenomenal concepts *do* represent phenomenal properties as absolutely intrinsic, especially given that <absolutely intrinsic> is such a sophisticated metaphysical concept. Second, even if our phenomenal concepts do represent phenomenal properties as absolutely intrinsic, we should not uncritically trust them. The very fact that the concepts represent something so metaphysically-loaded would be a strong reason to think that they might be inaccurate. In fact, Pereboom himself is quite willing to entertain this possibility.[21]

4.2.4 Arguments against materialism

A final sort of argument for (E) is to argue that materialism is false, conclude on this basis that there is an ontological gap between physical and phenomenal truths, and finally conclude that there is a corresponding *epistemic* gap between physical and phenomenal truths. To avoid circularity, the anti-materialist argument appealed to here must differ from the epistemic arguments that (E) is a premise of.

Of course, it would be pointless to argue in this way. If you have an independent argument against materialism, then there is no need to defend premise (E) of the epistemic arguments. You can just use that other argument to attack materialism directly.

Still, given this sort of argument for (E), the only way to show that there is no good argument for (E) is to show that there is no successful argument against materialism: if there *is* such an argument, then you can use it to show that (E) is

[21] The first half of Pereboom (2011) develops what he calls 'the qualitative inaccuracy thesis', which denies that our phenomenal concepts accurately represent the nature of phenomenal properties.

true. This is why I said above that is no good argument for (E) apart, perhaps, from appeals to considerations about conceptual analysis, *or from independent arguments against materialism*. As I said in chapter 1, I think that there is no successful argument against materialism, but my goal in this dissertation is not to show that *every* argument against materialism fails. My goal is only to show that the epistemic arguments fail.

The italicized restriction does not significantly weaken my defense of materialism. Either there is a successful independent argument against materialism or there is not. If there is such an argument, then it does not matter whether my attack on the epistemic arguments succeeds—materialism is false either way. If there is no such argument, then the restriction is irrelevant. This is because the unrestricted claim ("there is no good argument for (E) apart, perhaps, from appeals to considerations about conceptual analysis") follows from the restricted claim together with the claim that there is no successful independent argument against materialism.

4.3 Arguments against (E→O)

4.3.1 The phenomenal concept strategy

There are no arguments for (E) in the literature apart from the four that I just considered. Appeals to intuitions failed. The structure and dynamics argument either (i) appealed to considerations about phenomenal concepts, (ii) appealed to intuitions, or (iii) relied on controversial claims about what our phenomenal concepts represent and the accuracy of these representations. So, barring some new argument for (E), there is only one promising way to argue for (E) without making an independent anti-materialist argument: appeal to considerations

about phenomenal concepts, and use the Analysis Principle to infer from this that there is an epistemic gap. I will now argue that anti-materialists can defend (E→O) from two major objections only by *rejecting* this argument for (E) and denying the Analysis Principle.

The first objection comes from the PCS. As we saw in chapter 2, proponents of the PCS make arguments of the following form:

(1) *Attribution:* Phenomenal concepts are ϕ.

(2) *Explanation:* The fact that phenomenal concepts are ϕ explains the epistemic gap.

(3) *No evidence:* The fact that phenomenal concepts are ϕ is no evidence of an ontological gap.

(4) $\therefore$ The epistemic gap is no evidence of an ontological gap.

I argued in chapter 2 that (4) entails that, unless we have some independent reason to think either that there is an ontological gap or that there is no epistemic gap, we have no reason to accept (E→O). So proponents of the epistemic arguments must argue that the PCS fails, since they must claim that you can have reason to accept both (E) and (E→O) even if you do not already accept that there is an ontological gap.

If the conceptual considerations argument for (E) is sound, then the PCS succeeds. This can be seen as follows. When anti-materialists point to some feature of phenomenal concepts, ϕ, in arguing that there is an epistemic gap, materialists can point to ϕ to *explain why* there is an epistemic gap. Materialists can then plug that feature of phenomenal concepts into the above argument schema to get a defense of the PCS. So suppose that ϕ is, as in the above argument

for (E), the property of being non-analyzable.[22] Then we get the analyzability version of the PCS:

(A1) Phenomenal concepts are non-analyzable.

(A2) The fact that phenomenal concepts are non-analyzable explains the epistemic gap.

(A3) The fact that phenomenal concepts are non-analyzable is no evidence of an ontological gap.

(A4) ∴ The epistemic gap is no evidence of an ontological gap.

Anti-materialists who accept the above argument for (E) must accept (A1), since they themselves claim that phenomenal concepts are non-analyzable. They arguably must also accept (A2): if the fact that phenomenal concepts are non-analyzable entails that there is an epistemic gap (as the Analysis Principle claims), then the fact that phenomenal concepts are non-analyzable arguably *explains* the epistemic gap.[23] Alter (2015, pp. 6-8) even takes one of the main upshots of the structure and dynamics argument to be that the fact that phenomenal concepts are non-structural (roughly: non-analyzable, on his understanding of structure) explains the epistemic gap.

That just leaves (A3). Unlike (A1) and (A2), (A3) is not directly supported by anything anti-materialists say in defense of (E). The quick argument is to note that we have some non-analyzable concepts of uncontroversially physical things. Consider, for example, the concept of <life>. As Block and Stalnaker (1999, pp. 14-15) suggest, there is arguably no analysis of <life> into physical

[22] Remember that non-analyzable concepts are those that cannot be analyzed into *physical or functional* concepts. They need not be conceptually primitive.

[23] As I noted in chapter 2, here I am assuming merely that entailment is a defeasible guide to explanation, not that entailment and explanation always go hand-in-hand.

or functional concepts, but this is no evidence that every living thing is non-physical or that there is an ontological gap between the physical truths and truths about each living thing. Given this, it seems plausible that the mere fact that phenomenal concepts are non-analyzable is no evidence of an ontological gap between physical and phenomenal truths.

There is one loose end to tie up before moving on. Above I claimed that there are only four sorts of arguments for (E) in the literature. But it might be wondered whether the PCS opens the door to arguments for (E) that I did not consider above. After all, when proponents of the PCS point to some property of phenomenal concepts, ϕ, to *explain* the epistemic gap, anti-materialists can point to ϕ to *argue* that there *is* an epistemic gap.[24] This is basically the reverse of the move that I suggested that proponents of the PCS make against anti-materialists: where proponents of the PCS can use anti-materialist arguments for (E) to support the PCS, anti-materialists can use part of the PCS to argue for (E).

On reflection, though, this does not look like a promising option for anti-materialists. To make it work, they would have to argue that the relevant property of phenomenal concepts is evidence of an ontological gap, contrary to what proponents of the PCS claim. Otherwise the PCS succeeds. But this is implausible for any property of phenomenal concepts that proponents of the PCS actually appeal to. For example, the fact that phenomenal concepts are recognitional, quotational, or indexical (as on three leading versions of the PCS) would not be evidence of an ontological gap. I discuss this at more length in chapter 2. At any rate, no anti-materialists have taken up this line of argument.

[24]Thanks to an anonymous referee for raising this objection.

4.3.2 Other epistemic gaps

That, then, is one reason why anti-materialists must deny the Analysis Principle in order to defend (E→O): if the Analysis Principle is true, then the PCS succeeds. But there is another reason why anti-materialists must deny the Analysis Principle in order to defend (E→O).

(E→O) claims that one specific epistemic gap, the epistemic gap between physical and phenomenal truths, entails an ontological gap. But the motivation for (E→O) comes from the motivation for a more general principle connecting epistemic gaps to ontological gaps. The idea is that *any* epistemic gap (or at least, any epistemic gap relevantly similar to the one between physical and phenomenal truths) entails an ontological gap.[25] To defend (E→O), anti-materialists must argue that there are no such epistemic gaps that do not entail corresponding ontological gaps. So one way to argue against (E→O) is to argue that this more general principle is false.

Block and Stalnaker (1999) argue against (E→O) along these lines. They argue that there is an epistemic gap between the physical truths and truths about ordinary macroscopic objects (such as the truth that my glass has water in it).[26] Since there is clearly no *ontological* gap between the physical truths and truths about ordinary macroscopic objects, it follows that epistemic gaps do not in general entail ontological gaps. Of course, (E→O) might still be true, but it is hard to see why we should accept it if the more general principle connecting epistemic gaps to ontological gaps is false.

[25] For example, Chalmers (1996, 2010) does not claim merely that the conceivability of zombies entails their possibility. He claims that conceivability *always* entails possibility of a sort. (More precisely, he claims that ideal conceivability always entails primary possibility.)

[26] Again, remember that by 'physical truths' I mean '*narrowly* physical truths', where these are the truths that can be stated in the language of fundamental physics (see chapter 1). 'My glass has water in it' does not count as a physical truth in sense, since it contains terms like 'glass' and 'water'.

Why do Block and Stalnaker think that there is an epistemic gap between the physical truths and truths about ordinary macroscopic objects? Their main argument rests on considerations about conceptual analysis.[27] They claim that at least some of our concepts of macroscopic objects cannot be analyzed into physical concepts, and then infer from this (by tacit appeal to the Analysis Principle) that there is an epistemic gap between the physical truths and truths about at least some macroscopic objects. This argument should be familiar: it has the same structure as the conceptual considerations argument for (E) in §4.2.2 above.

Anti-materialists have two options in responding to Block and Stalnaker's argument. The first option is to argue that every concept of a macroscopic object *can* be analyzed into physical or functional concepts. This option is, I think, unpromising. It is plausible that *some* of our concepts of macroscopic objects (tables? cars? light bulbs?) can be analyzed into physical or functional concepts. But it is far less plausible that *every* concept of a macroscopic object can be so analyzed. For example, Block and Stalnaker give compelling reasons for thinking that there is no analysis of <life> into physical or functional concepts.

The second, more promising option is to deny the Analysis Principle. So anti-materialists can admit that there is no analysis of concepts like <life> into physical or functional concepts, but deny that it follows from this that there is an epistemic gap between the physical truths and truths involving such concepts. This, in fact, is how Chalmers and Jackson (2001, pp. 320-22) respond to Block and Stalnaker's argument. They present various putative counterexamples to the Analysis Principle. For example, they claim that knowing the description of a Gettier cases puts you in a position to know that the relevant agent lacks

[27] See Block and Stalnaker (1999, p. 13-16). They also make arguments the appeal to considerations about indexical truths (e.g., 'I am in Ithaca') and negative truths (e.g., 'there are no ghosts').

knowledge, even though there is no analysis of <knows>.[28]

So the only promising way to resist Block and Stalnaker's argument against (E→O) is to deny the Analysis Principle. Note, however, that both Block and Stalnaker's argument *against* (E→O) and the most promising anti-materialist argument *for* (E) rely on the Analysis Principle. So these arguments stand or fall together. If Chalmers and Jackson are right that the Analysis Principle is false, then this saves (E→O) only at the expense of (E). An anti-materialist cannot claim both that Block and Stalnaker go wrong in assuming the Analysis Principle, and then use the same principle to defend (E).

4.4 Two ways to be a materialist

So far my focus has been on problems that anti-materialists face in defending both (E) and (E→O), depending on whether the Analysis Principle is true. The flip-side of this is that different versions of materialism are plausible depending on whether the Analysis Principle is true. I will end this chapter by briefly highlighting one version of materialism that is plausible if the Analysis Principle is true, and introducing a different (and mostly neglected) version of materialism that is plausible if the Analysis Principle is false.

To start, suppose that the Analysis Principle is true. Then type-B materialists can use the Analysis Principle to defend the PCS and deny (E→O), as I discussed in §4.3.1. I developed the PCS at length in chapter 2, so I will not say more about it here.

Now suppose, more interestingly, that the Analysis Principle is false. This is to suppose that conceptual analysis is not necessary to close epistemic gaps. This

[28] See chapter 5 for much more on the Analysis Principle.

opens up a new and more promising way to be a type-A materialist and deny (E).

Traditionally, type-A materialists have been either analytic functionalists (who claim that phenomenal concepts can be analyzed into physical or functional concepts) or eliminativists (who deny that there are any positive phenomenal truths). Some philosophers have denied there could even be a version of type-A materialism that denies both analytic functionalism and eliminativism. For example, in his taxonomy of views of the metaphysics of consciousness, Chalmers (2003a) says that

> [o]nce we accept that the concept of consciousness is not itself a functional concept and that physical descriptions of the world are structural-dynamic descriptions, there is simply no conceptual room for it to be implied by a physical description (p. 122).[29]

But if the Analysis Principle is false, then this a mistake. It might be that analytic functionalism and eliminativism are false, but the physical truths *still* a priori entail the phenomenal truths.

This is important because the main motivation for rejecting type-A materialism has been the implausibility of analytic functionalism and eliminativism. For example, Chalmers (2003a) says that "[t]he obvious problem with type-A materialism is that it appears to deny the manifest" (p. 109). What he means is that it appears to deny the "manifest" truths that we are conscious and that phenomenal concepts are not physically or functionally analyzable. But if the Analysis Principle is false, then type-A materialists need not deny either of these

[29]Chalmers does briefly consider the possibility of a non-eliminativist, non-analytic functionalist form of type-A materialism, but then quickly goes on to to say that "in practice, there appears to be little room for such a view" and that there are "few examples of such views in practice" (p. 109). The few possible examples he mentions are basically slight variations on analytic functionalism.

claims. They can deny both analytic functionalism and eliminativism, and still deny that there is an epistemic gap between physical and phenomenal truths.[30]

So just as proponents of the epistemic arguments have a problem whether or not the Analysis Principle is true, materialists have a good option whether or not the Analysis Principle is true. If it is true, then they can accept the PCS and deny (E→O). If it is false, then they can accept a version of type-A materialism that is freed of implausible commitments to analytic functionalism and eliminativism, and deny (E).

4.5 Conclusion

I have argued that there is no good way for anti-materialists to defend both (E) and (E→O), short of appealing to an independent anti-materialist argument. If the Analysis Principle is true, then there is no good defense of (E→O). If it is false, then there is no good defense of (E)—apart, perhaps, from an independent anti-materialist argument. As I noted above, this does not show that (E) and (E→O) themselves are inconsistent. But if there is no good defense of both (E) and (E→O), then materialists do not need to accept them. So the Analysis Principle's truth or falsity affects *how* the epistemic arguments fail, not *whether* they fail.

Is the Analysis Principle true? As I have stressed, materialists do not need to answer this question in order to respond to the epistemic arguments. Still, the Analysis Principle is interesting in its own right, and it would be nice to know whether it is true. So in the next chapter I assess the Analysis Principle. I ultimately follow Chalmers and Jackson (2001) in rejecting it, but it worth seeing

[30]The only other defender of this form of type-A materialism that I know of is Boutel (2013). Note that he classifies the view as version of what Chalmers calls "type-C" materialism—his paper is even titled 'How to be a type-C physicalist'—but in fact the view fits Chalmers' definition of type-A materialism.

why it is false.

Chapter 5

Conceptual analysis and a priori entailment

5.1 Introduction

As we have seen, claims about epistemic gaps play a large role in debates over materialism. Anti-materialists claim that there is an epistemic gap between physical and phenomenal truths. Some materialists respond by claiming that there is no such epistemic gap.[1] Some admit that there is an epistemic gap between physical and phenomenal truths, but claim that there are also epistemic gaps between physical truths and truths about ordinary macroscopic objects.[2] Some admit that there is a *unique* epistemic gap between physical and phenomenal truths, but claim that it arises because of special features of phenomenal concepts.[3]

[1]For example, Lewis (1990), Dennett (1991), and Dretske (1995). See Chalmers (2003a) for more references.

[2]For example, Levine (1998, 2000), Block and Stalnaker (1999), Byrne (1999), and Schaffer (2017).

[3]For example, certain proponents of the PCS. ("Certain" and not "all" because some proponents of the PCS deny that the epistemic gap between physical and phenomenal truths is unique.)

How plausible these different claims are depends largely on what it takes for an epistemic gap to exist. One key question in assessing them is whether, and to what extent, conceptual analysis is necessary to close epistemic gaps. Some philosophers have thought that it is necessary, and so have accepted the Analysis Principle:

> **Analysis Principle:** If the concepts involved in one truth differ from and cannot be analyzed into the concepts involved in another truth, then there is an epistemic gap between the two truths.

These philosophers then go on to use the Analysis Principle to argue that certain epistemic gaps exist. Other philosophers have thought that conceptual analysis is not necessary to close epistemic gaps, and so have denied the Analysis Principle.[4] If this is right, then you cannot show that there is an epistemic gap between two truths simply by pointing to facts about conceptual analysis.

In this chapter I will review some arguments for and against the Analysis Principle and highlight its importance to various debates in the philosophy of mind. I ultimately follow Chalmers and Jackson (2001) in concluding that the Analysis Principle is false. This is bad for proponents of the analyzability version of the PCS. But as I argued in chapter 4, it is *also* bad for anti-materialists, since it undermines their best argument for the claim that there is an epistemic gap between physical and phenomenal truths.

5.2 Clarifying the principle

As it is stated above, the Analysis Principle is extremely general: it talks about epistemic gaps generally, without focusing on any particular sort of epistemic

See §2.1 above for many references.

[4] See §5.3 below for references.

gap (say, knowledge or conceivability gaps). Going forward, it will make things simpler if we narrow our focus a bit. So I will be assessing a restricted version of the Analysis Principle, which focuses on a priori entailment specifically. This more restricted principle says roughly that conceptual analysis is necessary for a priori entailment. Other discussions of the relation between conceptual analysis and epistemic gaps (for example, Chalmers and Jackson 2001) have also focused on a priori entailment.

This version of the Analysis Principle can be stated more precisely as follows:

(AP) *A* a priori entails *B* only if each concept involved in *B* either is the same as a concept involved in *A* or can be analyzed into the concepts involved in *A*.

There are many technical terms in (AP) to explain.

A concept is *involved* in a sentence iff an expression in the sentence either directly or indirectly expresses the concept. Direct and indirect expression of a concept is understood as follows. The expression 'square' directly expresses the concept <square>—it stands for precisely that concept. The expression 'square' *indirectly* expresses the concept <four> because, although it does not stand for <four>, <four> is directly expressed by part of an expression which is synonymous with 'square': 'figure with four sides of equal length, joined at right angles'. So the sentence 'squares exist' involves, among other concepts, <square> and <four>.[5]

A a priori entails B iff the material conditional '$A \rightarrow B$' is a priori.[6] A set Δ of sentences a priori entails *A* iff the conjunction of the sentences in Δ a priori

[5]If 'squares exist' did not count as involving <four>, then (AP) would entail, incorrectly, that 'if squares exist, then four-sided figures exist' is not a priori (since 'four-sided figures exist' involves the concept <four>). So it is important for the plausibility of (AP) that conceptual involvement is understood, as above, to include both direct and indirect expressions of a concept.

[6]Strictly speaking, there should be corner quotes, not ordinary quotation marks, around '$A \rightarrow B$'. Since it will not cause any confusion, I will continue to use ordinary quotation marks where, strictly speaking, corner quotes belong.

entails A. Something is *a priori* iff it can be known with justification independent of experience. For example, 'if I have a brother, then I have a sibling' is a priori because it can be known with justification independent of experience, and so 'I have a brother' a priori entails 'I have a sibling'.

A priori entailment is closely related to other epistemic notions, such as conceivability and scrutability. On certain conceptions of conceivability, '$A\&\sim B$' is conceivable iff A does not a priori entail B.[7] And on certain conceptions of scrutability, B is scrutable from A iff A a priori entails B.[8] So although I will be explicitly concerned only with a priori entailment, what I will say is also relevant to conceivability and scrutability.

I understand conceptual analysis as I did in chapter 2, in terms of the a priority of certain sentences. Again, the rough idea is this. The concept expressed by the predicate F is analyzable into the concepts expressed by the predicates $G_1 \ldots G_n$ iff a sentence of the form 'for all x, Fx iff _____' is a priori, where the blank is filled in only by logical expressions and the predicates $G_1 \ldots G_n$. So the concept <bachelor> is analyzable into the concepts <unmarried> and <male> because the following sentence is a priori: 'for all x, x is a bachelor iff x is an unmarried male'.[9]

[7] Chalmers (2002, p. 149) calls this 'negative ideal conceivability'.

[8] Chalmers (2012, p. 40) calls this 'a priori scrutability'.

[9] Some philosophers have a much broader understanding of conceptual analysis than this. For example, Jackson (1998) talks as if there can be conceptual analysis even if no sentence of this form is a priori. On Jackson's broad understanding of conceptual analysis, it is practically trivial that conceptual analysis is necessary for a priori entailment. (AP) is an interesting claim only on a more restricted understanding of conceptual analysis.

5.3 Why the principle matters

5.3.1 Arguments against materialism

The epistemic arguments start by claiming that the physical truths do not a priori entail the phenomenal truths,[10] and then infer from this that the physical truths do not necessitate or ground the phenomenal truths. In defending the crucial claim about a priori entailment, anti-materialists often rely on considerations about conceptual analysis. For example, recall Chalmers' (1996) argument that I quoted in chapter 4: "For consciousness to be [a priori] entailed by a set of physical facts, one would need some kind of analysis of the notion of consciousness... and there is no such analysis to be had" (p. 104).

If (AP) is false and conceptual analysis is not necessary for a priori entailment, then this argument is unsound: the physical truths might a priori entail the phenomenal truths even if there is no analysis of phenomenal concepts into physical concepts. Anti-materialists do have other arguments for the claim that there is no a priori entailment from the physical to phenomenal truths. For example, they also appeal to our intuitions that zombies are conceivable. But as I argued in chapter 4, these arguments do not work. So if (AP) false, then there is no good reason to accept the epistemic gap premise of the epistemic arguments.

5.3.2 The phenomenal concept strategy

Anti-materialists rely on (AP) to argue that there is no a priori entailment from the physical to phenomenal truths. In contrast, some materialists rely on (AP) to *explain why* there is no a priori entailment from the physical to phenomenal truths. This is a key part of some defenses of the phenomenal concept strategy

[10] Or that there is some other epistemic gap between them.

(PCS), which attempts to undermine the epistemic arguments by explaining the lack of a priori entailment from physical to phenomenal truths by appealing to facts about phenomenal concepts. I discussed the PCS at length in chapter 2.

Different versions of the PCS try to explain the lack of a priori entailment in different ways, and not all rely on (AP). But the versions that do rely on (AP) are some of the most popular ones. One version is the *analyzability* version, which tries to explain the lack of a priori entailment by claiming simply that phenomenal concepts cannot be analyzed into physical concepts.[11] If (AP) is false, then this explanation presumably fails: if there can be a priori entailment without conceptual analysis, then it is hard to see why the mere fact that phenomenal concepts cannot be analyzed into physical concepts would explain the lack of a priori entailment from physical to phenomenal truths.

5.3.3 Scrutability theses

One truth, B, is (a priori) scrutable from another truth, A, iff A a priori entails B. In *Constructing the World* (2012), Chalmers defends various scrutability theses, which claim that *every* truth is scrutable from a certain relatively small set of truths.[12] For example, one scrutability thesis that Chalmers defends is *fundamental scrutability*: the claim that every truth is scrutable from the fundamental truths. If (AP) is true, then fundamental scrutability is implausible. This is because it is implausible that, for any truth, the concepts involved it it can be analyzed into the concepts involved in the fundamental truths.[13] So it is crucial to Chalmers'

[11] This version of the PCS is conditionally defended by me in chapter 2. See Levine (1998, 2001) for a similar defense of the PCS. As I noted in chapter 2, most defenders of the recognitional and quotational versions of the PCS also appeal to (AP).

[12] Chalmers and Jackson (2001) defend a closely related view.

[13] For example, it is implausible that there is any analysis of <knows> (and a fortiori, it is implausible that there is any analysis of <knows> in terms of the concepts involved in fundamental truths), even though no truth involving <knows> is fundamental. See §5.5 below

project that (AP) is false.

Scrutability theses are interesting in their own right, but they also play an indirect role in arguments against materialism. This is because, as Levine (2014) suggests, the epistemic arguments are arguably committed to something like fundamental scrutability.[14] After all, if the fundamental truths do not a priori entail every truth, then it is hard to see why the mere fact that the physical truths do not a priori entail the phenomenal truths would show that there are fundamental non-physical truths. So one way of rebutting the epistemic arguments is to argue that fundamental scrutability is false.[15]

5.4 Motivation for the principle

(AP) is often assumed, but rarely defended. As best I can tell, its proponents often tacitly assume what I call 'the logical derivation model of a priori of entailment'. According to this model, A a priori entails B only if either B can be logically derived from A, or B is the result of substituting synonyms for synonyms in a sentence that can be logically derived from A. In other words, A a priori entails B only if '$A \rightarrow B$' is either a logical truth or synonymous with a logical truth.[16] It is clear that this is *sufficient* for a priori entailment. The logical derivation modal of a priori entailment claims more controversially that it is *necessary* for a priori entailment.

The argument from the logical derivation model of a priori entailment to (AP)

for discussion of a closely related point.

[14] Levine calls fundamental scrutability the 'A Priori Entailment' (AE) thesis.

[15] This is similar Block and Stalnaker's (1999) response to the epistemic arguments, which I considered in chapter 4.

[16] Boghossian (1997) calls such truths 'Frege analytic'. To get a sense of the distinction between logical truths and Frege analytic truths, consider: 'all bachelors are unmarried men' is not a logical truth (since it is of the form '$\forall x(Fx \rightarrow Gx)$'), but it is Frege analytic (since it synonymous with the logical truth 'all unmarried men are unmarried men').

is relatively straightforward. Here is the basic idea (setting aside a complication raised in the next paragraph). If B involves a concept which is not involved in A and cannot be analyzed into the concepts involved in A, then '$A \rightarrow B$' is neither a logical truth nor synonymous with a logical truth. So, given the logical derivation model of a priori entailment, if B involves a concept which is not involved in A and cannot be analyzed into the concepts involved in A, then A does not a priori entail B. Contrapositively: A a priori entails B only if each concept involved in B either is the same as a concept involved in A or can be analyzed into the concepts involved in A.

A complication arises from the fact that, for some A and B, '$A \rightarrow B$' *is* a logical truth or synonymous with a logical truth even though there is no analysis of the concepts involved in A into the concepts involved in B. For example, suppose that B is itself a logical truth. Then for any A, '$A \rightarrow B$' is also a logical truth, regardless of whether there is an analysis of the concepts involved in B into the concepts involved in A. (Note that there is the same sort of counterexample to Hume's claim that you cannot derive an 'ought' from an 'is'. For example, 'either you ought to ϕ or it's not the case that you ought to ϕ' is a logical truth, and so trivially derivable from any sentence.)

This is a well-known and surprisingly difficult problem.[17] One quick fix is to restrict B to atomic sentences that are not logical truths.[18] (A sentence is atomic iff it does not have sentences as proper parts. For example, 'A&B' is not an atomic sentence, since it has the sentences A and B as proper parts.) So we can claim that

[17] See Restall and Russell (2010) and Stoljar (2006, p. 27) for discussion. The fix that I suggest in the main text is similar to Stoljar's suggested fix.

[18] It is not enough to simply restrict B to non-logical truths. For example, 'if I have a finger, then either I have a finger or I am phenomenally conscious' is a logical truth even though (i) 'either I have a finger or I am phenomenally conscious' is not a logical truth and (ii) there is no analysis of <phenomenally conscious> into the concepts involved in 'I have a finger'. And it is not enough to simply restrict B to atomic sentences, because some atomic sentences are logical truths (for example, 'TF = TF').

for any sentence A and any *atomic*, *non-logical* sentence B, if B involves a concept which is not involved in A and cannot be analyzed into the concepts involved in A, then '$A \rightarrow B$' is neither a logical truth nor synonymous with a logical truth. This does not entail (AP), but it does entail (AP*):

(AP*) If B is a non-logical atomic truth, then A a priori entails B only if each concept involved in B either is the same as a concept involved in A or can be analyzed into the concepts involved in A.

(AP*) is weaker than (AP), but even (AP*) is relevant to debates over materialism, the PCS, and scrutability theses. For example, anti-materialists can use (AP*) to argue that non-logical atomic phenomenal truths, such as 'TF is phenomenally conscious' are not a priori entailed by the physical truths. Proponents of the PCS can try to use (AP*) to explain why such truths are not a priori entailed by the physical truths. And proponents of scrutability theses will want to argue that even (AP*) is false, since they claim that even non-logical atomic truths (for example, 'TF is in Ithaca') are scrutable from relatively small sets of truths.

So if the logical derivation model of a priori entailment is right, then there is a compelling case to be made for (AP) or something close enough to it. The key question, then, is whether the logical derivation model of a priori entailment is right. In its favor, it makes clear how a priori entailment is possible. On the face of it, it is puzzling how one truth could a priori entail another one if there is no logical or analytic connection between the two truths. But this is far from a knock-down argument for the model. And as we will see, there are some compelling counterexamples to it. I discuss these in the next section, when considering counterexamples to (AP) itself.

5.5 Counterexamples to the principle

(AP) is false. There are two sorts of counterexamples to it, which are also counterexamples to the logical derivation model of a priori entailment.

The first sort of counterexample to (AP) is structurally similar to what we saw in the previous section. Suppose that B is a priori. Then for any A, A a priori entails B, regardless of whether there is an analysis of the concepts involved in B into the concepts involved in A. The quick fix is to give up (AP) in favor of (AP*), which restricts B to non-logical atomic sentences.[19] As I noted above, even the more modest principle (AP*) could play an important role in debates over materialism, the PCS, and scrutability theses. So I do not think that proponents of (AP) should be seriously worried about this sort of counterexample to it.

The second sort of counterexample to (AP) is more serious, since there is no obvious way to modify (AP) to accommodate it. The specific counterexample that I will consider is inspired by Chalmers and Jackson (2001), but I have changed the details of it. Let J be the conjunction of sentences in the following description:

> Jones is a reliable mathematician who almost never makes mistakes. One day he works out a correct proof for the Pythagorean theorem, and on the basis of this proof he forms the belief that the theorem is true. Though the proof is simple, he double- and triple-checks his work, and finds no mistakes. He then gets a colleague to check the proof, and the college confirms that there are no mistakes. Both Jones and his college are thinking clearly. Jones believes with justification everything stated in this description.

Let K be the sentence, 'Jones knows that the Pythagorean theorem is true'. J a

[19] This avoids the counterexample only if B is either non-atomic or non-logical. If there are a priori sentences that are both atomic and non-logical, then B must be further restricted to a posteriori sentences.

priori entails *K*. But there is no analysis of <knows> into the concepts involved in *J*.[20]

In response to this counterexample, defenders of (AP) have two options. First, they could deny that *J* a priori entails *K*. For example, they could argue that the inference from *J* to *K* is really induction or inference to the best explanation and not a priori.[21] Second, they could claim that <knows> *is* analyzable into the concepts involved in *J*—it is just an extremely complicated and non-obvious analysis. Neither option is promising.

Against the first option: If the inference from *J* to *K* is not a priori, then there is a conceivable situation in which '*J*&∼*K*' is true. For example, the inference from 'the ball was dropped' to 'the ball fell' is merely inductive and not a priori, and so there is a conceivable situation in which 'the ball was dropped but didn't fall' is true. But there is no conceivable situation in which '*J*&∼*K*' is true: try as I might, I cannot conceive of a situation in which Jones fails to know that the theorem is true even though he is a reliable mathematician, his proof is correct, he is thinking clearly, and so on. So the inference from *J* to *K* is a priori.[22]

Against the second option: As Chalmers and Jackson (2001, p. 321) note, there have been decades of failed attempts to analyze <knows>. This long history of failed analyses makes it plausible that there is no analysis of <knows> at all, let alone an analysis into the concepts involved in *J*.[23] On top of this, even if there

[20]The counterexample presented by Chalmers and Jackson involves a Gettier case—a case in which the agent lacks knowledge despite having justified true belief (Gettier 1963). I have presented a case in which the agent *has* knowledge because I want a counterexample not only to (AP) but also to (AP*). From the description of a Gettier case, we can a priori infer that the agent does not have knowledge. But this is not a counterexample to (AP*) because the sentence 'the agent does not have knowledge' is non-atomic (since it is a negation), and so (AP*) does not apply to it. In contrast, the sentence 'the agent has knowledge' is atomic, and so (AP*) applies to it.

[21]A different, more radical option is to be an a priori skeptic: to deny that *any* sentences are a priori.

[22]Chalmers and Jackson (2001) do not argue in this way, but Chalmers often uses this strategy to argue for a priori entailment claims. For example, see his (1996) and (2012).

[23]See Ichikawa and Steup (2017) for an overview of attempts to analyze <knows>.

turns out to be some complicated analysis of <knows>, we are not relying on that analysis when we infer *K* from *J* since we have no idea what the analysis is. So even if this case is strictly speaking compatible with (AP) because there turns out to be an analysis of <knows>, it still gives us reason to doubt that conceptual analysis is necessary for a priori entailment.

Finally, it is worth noting that there are many other counterexamples to (AP). For example, 'something is red' a priori entails 'something is colored', even though there is no analysis of <colored> in into <red>.[24] So even if a defender of (AP) can somehow show that *J* does not a priori entail *K* or that there is an analysis of <knows>, that is not enough to save (AP). They must work through every other putative counterexample to (AP), such as the red/colored example just mentioned, and show the same thing about it. And it is extremely implausible that they will be able to do this in every case. Maybe they can do it in one or two cases, but only one counterexample needs to succeed for (AP) to be false.

5.6 Approximate analyses

(AP) is false. There is, however, weakened version of the analysis principle which tries to avoid the counterexamples in the previous section. This principle claims not that having a perfect, counterexample-free analysis of a concept is necessary for a priori entailment to truths involving the concept. Instead, it claims merely that having at least an *approximate* analysis of the concept is necessary for a priori entailment to truths involving the concept. Call this 'the Approximate Analysis

[24] I have seen this example in many different places, not always aimed at attacking (AP). Chalmers (2014, p. 800) mentions it as a counterexample to (AP). It also comes up in Fodor (1998, pp. 109-10) and Williamson (2001, pp. 23-4) I do not know what its original source is.

Principle'. Chalmers (1996) seems to allude to this sort of principle in a few places (for example, pp. 53-4 and pp. 80-1).

I will clarify and assess the principle below, but it is worth first explaining why the principle matters. Most notably, even the Approximate Analysis Principle can play an important role in arguments against materialism. Anti-materialists can claim that there is not even an approximate analysis of phenomenal concepts into physical or functional concepts, and then use the Approximate Analysis Principle to conclude that the physical truths do not a priori entail physical truths.

Importantly for anti-materialists, the Approximate Analysis Principle does not obviously entail that epistemic gaps are rampant. For example, although there is no perfect, counterexample-free analysis of <life> into physical or functional concepts, there is arguably at least an *approximate* analysis of it: following (Chalmers 1996, p. 53), we can say that something is alive iff it reproduces, adapts, and metabolizes.[25] So the principle does not entail that there is an epistemic gap between physical truths and truths involving <life>. This is what anti-materialists want: if epistemic gaps entail ontological gaps, as anti-materialists claim, then we should not find them where we know there are no ontological gaps.

There are, however, two difficult challenges for proponents of the Approximate Analysis Principle. The first is to explain what exactly it means for there to be an approximate analysis of a concept. The second is to motivate and defend the principle. I do not think that either challenge can be met.

[25]Even this analysis is not in narrowly physical terms, understood as in chapter 1 (that is, the terms of fundamental physics). This is a problem for Chalmers, since it is not obvious that there *is* an approximate analysis in narrowly physical terms, even if there is an approximate analysis in broadly physical terms. But I will set this problem aside here.

5.6.1 Clarifying the weakened principle

A little more precisely, the Approximate Analysis Principle can be stated as follows:

(AAP) *A* a priori entails *B* only if each concept involved in *B* either is the same as a concept involved in *A* or can be *at least approximately* analyzed into the concepts involved in *A*.

But what exactly is an approximate analysis of a concept? Roughly, is that it is an analysis that is at least mostly, but not necessarily perfectly, extensionally correct. For example, the JTB (justified true belief) analysis of <knows> is an approximate analysis of it: in most cases someone knows iff they have JTB, but there are edge cases where someone fails to know even though they have JTB. Another example is the approximate analysis of <life> from above.

This is still, however, only a rough account of what approximate conceptual analysis comes to. Above I characterized ordinary, non-approximate conceptual analysis in terms of a priority: *F* is analyzable into the concepts expressed by the predicates $G_1 \ldots G_n$ iff a sentence of the form 'for all *x*, *Fx* iff _____' is a priori, where the blank is filled in only by logical expressions and the predicates $G_1 \ldots G_n$. But it is unclear how this definition can be adapted to approximate conceptual analysis. After all, it is unclear what it would mean to say that a sentence is *approximately* a priori.

To avoid this problem, we could try saying that in approximate analyses it is not that something is approximately a priori. Rather, it is (precisely) a priori that something is approximately true. So *F* is approximately conceptually analyzable into the concepts expressed by the predicates $G_1 \ldots G_n$ iff a sentence of the form '*it is at least approximately true that* for all *x*, *Fx* iff _____' is a priori, where the blank is filled in only by logical expressions and the predicates $G_1 \ldots G_n$. But this

runs into other problems. For example, it is not a priori that the JTB analysis of <knows> is approximately correct. After all, for all I know a priori, Gettier cases are commonplace.

A final try is to drop the a priori condition. So F is approximately conceptually analyzable into the concepts expressed by the predicates $G_1 \ldots G_n$ iff a sentence of the form 'it is at least approximately true that for all x, Fx iff _____' is a *true* (not necessarily a priori), where the blank is filled in only by logical expressions and the predicates $G_1 \ldots G_n$. But this, unsurprisingly, runs into other problems. For example, it has the consequence that <water> is approximately conceptually analyzable as <H_2O>, and that this was true even before it was discovered that water is H_2O. This is just not the sort of analysis that philosophers have in mind in when they talk about *conceptual* analysis—especially not if there is supposed to be an interesting connection between conceptual analysis and a priori entailment, as proponents of (AAP) claim.

It is unclear to me how else the notion of an approximate analysis can be clarified. And if it cannot be clarified, then (AAP) is objectionably unclear.

5.6.2 Assessing the weakened principle

Suppose, however, that (AAP) can be adequately clarified. There is still the question of whether it is true. In its favor, (AAP) seems to avoid at least some of the counterexamples to (AP). For example, although there is no counterexample-free analysis of <knows>, the JTB analysis is at least approximately correct. Since there is at least an approximate analysis of <knows> into the terms of J, (AAP) allows that J a priori entails K.

Still, there are nearby cases that put significant pressure on (APP), even if they are not outright counterexamples to it. For example, in their argument against

the original (AP), Chalmers and Jackson (2001) discuss a a Gettier case: a case in which the agent lacks knowledge despite having JTB. This is a counterexample to (AP) because the description of the Gettier case, *G*, a priori entails that the agent lacks knowledge, $\sim K$, even though there is no analysis of <knows>. This is not obviously a counterexample to (AAP), because there is at least an *approximate* analysis of <knows> in terms of JTB.

But this case still casts doubt on (AAP) because the approximate analysis of <knows> plays no role in someone's inferring $\sim K$ from *G*. After all, this is precisely one of the cases in which the approximate analysis of <knows> as JTB gives the wrong result—it entails, incorrectly, that the agent knows the relevant truth. That is the whole point of Gettier's example. But having an analysis of <knows> that entails the *wrong thing* about *G* cannot help you infer $\sim K$ (that the agent does *not* know the relevant truth) from *G*. So there is no reason to think that there being even an approximate analysis of <knows> is necessary for *G* to priori entail $\sim K$. Given this, I think that (AAP) is implausible.

5.7 Conclusion

The Analysis Principle is false, and the Approximate Analysis Principle is unclear and at best implausible. So if proponents of the PCS want to explain why there is an epistemic gap, they cannot appeal to either of these principles. And if anti-materialists want to argue that there is an epistemic gap, then they also cannot appeal to these principles. If what I say in chapter 4 is right, then this is basically a wash: the most promising materialistic explanation of the epistemic gap (assuming there is such a gap) fails, but the most promising argument that the epistemic gap even exists *also* fails.

Appendix A

The truth of schmenomenal beliefs

A.1 Introduction

In chapter 3, I argued that proponents of the PCS need not explain the truth of our phenomenal beliefs in order to defend the PCS from Chalmers' dilemma. Given this, I defended a narrow correspondence thesis which claimed that zombies face an epistemic gap between physical truths and their schmenomenal beliefs, but which allowed that their schmenomenal beliefs might be false. I suggested, however, that a reasonable case can be made that at least some schmenomenal beliefs are true.[1] Call this 'the truth thesis'. In this appendix, I will defend the truth thesis.

The payoff is that the truth thesis can play a key role in an objection to the epistemic arguments that is independent of the PCS. This objection is more ambitious than the PCS in that it aims to show not merely that the epistemic arguments fail, but also more strongly that they are *unsound*. (Recall from §2.3

[1] In fact, I think there is a case to be made that *most* schmenomenal beliefs are true, but for the purposes of this appendix (see the next paragraph in the main text) it is enough to show that some schmenomenal beliefs are true.

that the PCS does not aim to show that the epistemic arguments are unsound.)

This appendix is more speculative than the rest of this dissertation. I do not claim that there is a knock-down argument for the truth thesis. Still, given the possibly high payoff if it is true, it is worth at least making the case for it.

As in chapter 3, for ease of presentation I will not bother qualifying talk of zombies with a conceivability operator. So I will say things like "zombies are X" when I mean "*it is conceivable* that zombies are X". My claim, then, is not that zombies exist and have true schmenomenal beliefs (which would be inconsistent with materialism), but rather that it is conceivable that they exist and have true schmenomenal beliefs.

A.2 The initial argument

My argument for the truth thesis relies on the assumption that zombies have concepts, and that their concepts work pretty much how our concepts work.

As I noted in chapter 3, some philosophers would deny that this is conceivable. For example, McGinn (1988) and Searle (1992) believe that consciousness is necessary for intentionality, including the grasping of concepts and the holding of beliefs. Searle calls this 'the Connection Principle'. If this principle is a priori—if it is a priori something can grasp concepts or have beliefs only if it is conscious—then it is inconceivable that zombies have concepts or beliefs. So I will be assuming that the Connection Principle is at least conceivably false, and so it is conceivable that zombies have concepts and beliefs even though they lack consciousness.

My general strategy in defending the truth thesis is to argue that if zombies' schmenomenal concepts work pretty much how our phenomenal concepts work,

then their concepts refer to physical or functional states that zombies are often actually in (as opposed to, say, failing to refer or referring to phenomenal states). And their schmenomenal concepts *should* work pretty much how our phenomenal concepts work, since zombies are functional duplicates of us. If all this is right, then at least some of zombies' schmenomenal beliefs are true.

Here I will focus on two popular accounts of phenomenal concepts: the recognitional and quotational accounts. I discussed these accounts in previous chapters, so I will only briefly summarize them here. According to the *recognitional account*, phenomenal concepts are recognitional concepts: concepts that refer to whatever typically triggers their applications.[2] According to the *quotational account*, phenomenal concepts are quotational concepts: concepts that refer to whatever they are constituted by. My claim, then, is that if schmenomenal concepts are either recognitional or quotational, then they refer to physical or functional states that zombies are often in.

Recognitional account: Suppose that phenomenal concepts are recognitional concepts, and so schmenomenal concepts are also recognitional concepts. A recognitional concept refers to whatever typically triggers its applications. So schmenomenal concepts refer to whatever typically triggers their applications—presumably certain brain states or functional states that zombies are actually in. If this is right, then at least some schmenomenal beliefs are true. (Shoemaker 1999, p. 411 briefly considers this possibility.)

It might be objected that phenomenal concepts, even if they are recognitional, do not refer simply to whatever typically triggers their applications. Rather, they refer to whatever *phenomenal states* typically trigger their applications. If the

[2]I am working with a simplified account of recognitional concepts. An actual account would have to be more complicated to account for, among other things, the fact that there will often be more than one thing that could count as triggering the application of a recognitional concept. For example, both an individual plant and a type of plant could count as triggering the application of the same recognitional concept.

applications of a phenomenal concept are typically triggered by something that is not a phenomenal state, then it fails to refer. So schmenomenal concepts fail to refer, since they are not typically triggered by phenomenal states. Analogy: we might have a recognitional concept, <triffid>, that refers to whatever *plant* typically triggers its applications. If it turns out that applications of <triffid> are typically triggered by a mirage, then <triffid> fails to refer and does not refer to the mirage.

In response, I grant for the sake of argument that our phenomenal concepts work as suggested in the previous paragraph. So it is built into them that they refer to what phenomenal states typically trigger their applications. Note, however, that the schmenomenal concepts of zombies will work slightly differently, since zombies have no concept of phenomenal states (just as our counterparts on Twin Earth have no concept of water). Instead, it will be built into zombies' schmenomenal concepts that they refer to whatever *schmenomenal* states typically trigger their applications. And there *are* schmenomenal states in the zombie world to trigger the applications of schmenomenal concepts.

Here is an analogy that might help explain this point. Suppose that we have a recognitional concept, W, that refers whatever specific glass of water typically triggers its applications. Our counterparts on Twin Earth will have a corresponding concept, TW. But it would be a mistake to claim that since there are no glasses of water on Twin Earth, TW must be an empty concept. TW will not refer to whatever glass of *water* (i.e., H_2O) typically triggers its applications. Instead, it will refer to whatever glass of *twater* (i.e., XYZ) typically triggers its applications. So the content of the restrictions built into a recognitional concept can shift between corresponding concepts.

It might then be objected that a different restriction is built into phenom-

enal concepts, and that the content of this restriction does not shift between corresponding concepts. For example, maybe it is built into our phenomenal concepts that they refer to whatever *simple* states typically trigger their applications. It is plausible that the content of the concept that corresponds to our concept <simple> does not differ in the zombie world. If this is right, then zombies' schmenomenal concepts are all empty: the applications of their schmenomenal concepts are triggered by *complex* brain states or functional states, not by anything simple.

In response, I grant that if our phenomenal concepts work as suggested in the previous paragraph, then zombies' schmenomenal concepts are all empty. But I deny that our phenomenal concepts work as suggested in the previous paragraph. If our phenomenal concepts worked as suggested in the previous paragraph, then they might *also* be empty, since their applications might not be triggered by anything simple. After all, if materialism is true, then phenomenal states are complex brain states or functional states. So we should reject any account of phenomenal concepts that builds a simplicity condition into them. Even if materialism is false, it is not *analytically* false.

More generally, it is possible to give accounts of our phenomenal concepts such that that, if zombies' schmenomenal concepts work in the same way, then their schmenomenal concepts are empty. But many of these accounts entail that our phenomenal concepts are empty, too, unless materialism is false. Since it is not analytic that materialism is false, these these accounts of phenomenal concepts should be rejected.

Quotational account: The argument here is pretty much the same as the argument for the recognitional account. Suppose that phenomenal concepts are quotational concepts, and so schmenomenal concepts are also quotational concepts.

A quotational concept refers to whatever it is constituted by. So schmenomenal concepts refer to whatever they are constituted by by—again, presumably certain brain states or functional states that zombies are actually in. If this is right, then at least some schmenomenal beliefs are true.[3]

It might be objected that zombies could not even have quotational concepts, or that quotational concepts can refer only to phenomenal states (and so zombies could have at most *empty* quotational concepts). In response, I deny this. On the best developed account of quotational concepts, there is little reason to think that quotational concepts can be had only by conscious agents or that they can refer only to phenomenal states.

On Balog's (2012a) account of quotational concepts, they are the mental analogue of quote terms. Just as you form a quote term by putting quotation marks around an expression, she thinks that you form a quotational concept by putting the mental equivalent of quotation marks around a mental state. Ordinary quotation marks can be fully characterized by their conceptual role—specifically, the role they play in disquotation. Similarity, Balog thinks that mental quotation marks can be fully characterized by their conceptual role. Importantly, nothing about this conceptual role requires a connection to phenomenal states.

To see this, consider more fully the parallel that Balog draws between quote terms and quotational concepts. Everyone who understand quote terms is disposed to accept all instances of the following sentence schema:

'x' refers to x

where x is a term. Balog suggests that the conceptual role of quotational concepts can be characterized in an analogous way. Everyone who grasps quotational concepts is disposed to accept all instances of the following thought schema:

[3]Papineau (2007, pp. 140-3) seems to endorse this argument.

__m__ REFERS-TO __m__

where __m__ is a mental state—not necessarily a *conscious* mental state[4]—and the asterisks are the mental counterpart of quotation marks. I am unsure whether the exact details of Balog's suggestion are right, but I think that it makes it plausible that quotational concepts are not necessarily connected in any way to phenomenal states.

That, then, is my argument for the truth thesis—the claim that some schmenomenal beliefs are true. The main premises it relies on are (i) that zombies' schmenomenal concepts are recognitional or quotational, and (ii) if so, then they refer to brain states or functional states of zombies. Both, I think, are plausible, but I have not made anything like a knock-down argument for them.

Balog (1999, p. 519) makes a different argument for the truth thesis. It goes like this. Zombies form a linguistic community. In general, we should try to interpret the utterances of members of a linguistic community in a way that makes most of them come out as true. So, in particular, we should try to interpret the utterances of zombies in a way that makes most of them (including utterances like 'I am in pain') come out as true. The best way to do this is to interpret 'pain' and similar terms, as used by zombies, as referring to physical or functional states that zombies are often in. On this interpretation, most schmenomenal beliefs are true because zombies are often in the relevant physical or functional states.

I do not think that Balog's argument works. The problem is that it applies the principle of charity to members of a merely possible (or merely conceivable!) linguistic community. And it is not obvious that the principle of charity should

[4]Balog (2012a) actually says 'experience' where I say 'mental state', but nothing in her account requires this restriction to experiences. In fact, her analogy suggests that there should be no such restriction. After all, there is no restriction on which expressions can be (linguistically) quoted, so why should there be any restriction on which mental states can be mentally quoted?

be applied to members of merely possible linguistic communities. Arguably, the reason why we should interpret actual speakers' utterances charitably is that in the *actual* world, people are usually not radically deluded. So an interpretation of their utterances on which they *are* radically deluded is probably wrong. But when we consider merely possible speakers, the same is not true: merely possible speakers, such as zombies, might well be radically deluded. All sorts of crazy stuff happens in merely possible worlds.

A.3 Chalmers' objection

Chalmers (2007) argues against the truth thesis.

First, some terminology. Distinguish between (phenomenal) realists and (phenomenal) eliminativists.[5] Realists believe that (i) phenomenal concepts are not functionally analyzable and (ii) we are phenomenally conscious. Eliminativists deny realism: they believe that either (i) phenomenal concepts are functionally analyzable or (ii) we are not phenomenally conscious. Note that realism entails nothing about whether phenomenal states themselves are functional states. It just entails that if they are, then this does not follow trivially from our phenomenal concepts.[6]

Now consider the zombie counterparts of realists and eliminativists—*zombie* realists and *zombie* eliminativists. Zombie realists believe that (i) schmenomenal concepts are not functionally analyzable and (ii) zombies are schmenomenally conscious. Zombie eliminativists believe that either (i) schmenomenal concepts are functionally analyzable or (ii) zombies are not schmenomenally conscious.

[5] I will omit the 'phenomenal' qualification in what follows. The distinction is from Chalmers (2003b).

[6] In Block's (1996) terminology, this means that realism is consistent with psychofunctionalism. It is just inconsistent with *analytic* functionalism.

Schmenomenal consciousness is whatever property or state, if any, that zombies refer to when they say 'phenomenal consciousness'.

Chalmers' argument then goes like this. In the *actual* world, realism is true and eliminativism is false. So realists are right and eliminativists are wrong. This much is common ground for anti-materialists and proponents of the PCS. But in the *zombie* world, things are different. Zombie realists are wrong and zombie eliminativists are right. So the schmenomenal beliefs of zombies are false (2007, pp. 177-8).

The key claim is that zombie eliminativists are right. There are a couple of ways to argue for this claim. Here is one natural but confused argument. (Note that Chalmers does not make this argument.) All you need to do to make an (actual-world) eliminativist right is to take away phenomenal consciousness from the actual world. And a zombie world is precisely that—a world identical to the actual world, except that phenomenal consciousness has been taken away. So eliminativism is true in the zombie world. So zombie eliminativists are right.

This argument is confused because it assumes falsely that zombie eliminativists believe the same thing as actual-world eliminativists. It is true that eliminativism, as believed by *actual-world* eliminativists, is true in the zombie world. But it does not follow from this that zombie eliminativists are right, since their beliefs differ from actual-world eliminativists. Again, zombie eliminativists believe that either (i) schmenomenal concepts are functionally analyzable or (ii) zombies are not schmenomenally conscious. Neither (i) nor (ii) is obviously true in the zombie world.

Chalmers has a different argument which aims to show that zombie eliminativists are right. He has us imagine the following dialogue between a zombie realist and a zombie eliminativist:

> Zombie Eliminativist: "There's no such thing as phenomenal consciousness."
>
> Zombie Realist: "Yes, there is."
>
> Zombie Eliminativist: "We are conscious insofar as 'consciousness' is a functional concept, but we are not conscious in any further sense."
>
> Zombie Realist: "No, we are conscious in a sense that is not functionally analyzable." (2007, p. 178)

Intuitively, the zombie eliminativist is right and the zombie realist is wrong.

But for the reasons I discussed above, I think that the zombie realist is right and the zombie eliminativist is wrong. Recall that zombie realism is true iff (i) schmenomenal concepts cannot be analyzed into physical or functional concepts and (ii) zombies are schmenomenally conscious. So to show that the zombie realist is right, I just need to show that (i) and (ii) are true in the zombie world. I defended both these claims above (§§3.5.3, A.2), but it is worth briefly going through the arguments again.

In support of (i): Zombies are functional duplicates of us, and so phenomenal and schmenomenal concepts play the same conceptual role. Phenomenal concepts cannot be analyzed into physical or functional concepts. So schmenomenal concepts cannot be analyzed into physical or functional concepts, either—otherwise phenomenal and schmenomenal concepts would play different conceptual roles. In support of (ii): Schmenomenal concepts are either recognitional or quotational. In either case, they refer to physical or functional states that zombies are actually in, and so zombies are schmenomenally conscious.[7]

[7] This quick summary glosses over some possible objections (that there is a descriptive element to recognitional concepts, that zombies could not even have quotational concepts) that I addressed above, in §A.2.

It might be objected that when the zombie realist says "we are conscious in a sense that is not functionally analyzable", he means in part that schmonsciousness is not a physical or functional state. And if schmonsciousness is not a physical or functional state, then zombies are not schmonscious. After all, in the zombie world there are only physical and functional states.

But this is not what the zombie realist means. He is not making a claim about schmonsciousness itself. He is making a claim about the *concept* of schmonsciousness: that it is not functionally analyzable. As I noted above, this is consistent with schmonsciousness being a physical or functional state. Neither actual-world realists nor zombie realists are committed to the falsity of functionalism. They are just committed to the view that functionalism, if true, does not follow trivially from the concepts of consciousness or schmonsciousness.

A.4 Connection to Balog's "zombie refutation"

If what I have said up to this point is right, then the truth thesis is true: at least some of zombies' schmenomenal beliefs are true. Why does this matter? Some proponents of the seem to PCS think that Chalmers' dilemma succeeds in refuting the PCS if all of zombies' schmenomenal beliefs are false, and so the truth thesis is necessary to save the PCS from Chalmers' dilemma.[8] But in chapter 3, I argued that this is wrong: proponents of the PCS can parry Chalmers' dilemma even if all of zombies' schmenomenal beliefs are false.

Instead, I think the truth thesis matters because of its role in a *different* argument which aims to undermine the epistemic arguments. This is a point that Balog (1999) makes in a paper written before Chalmers (2007) posed his dilemma

[8]Most notably, Carruthers and Veillet (2007) and Papineau (2007).

to proponents of PCS. She calls her argument, which relies on the claim that some of zombies' schmenomenal beliefs are true, 'the zombie refutation'. I will end this appendix by discussing the zombie refutation. As I noted above, the zombie refutation is more ambitions than the PCS in that it aims to show not just that the standard materialist arguments fail, but that they are unsound.

The zombie refutation goes like this. Zombies face an epistemic gap that corresponds to the epistemic gap that we face, as the correspondence thesis (defended in chapter 3) claims. And some of their schmenomenal beliefs are true, as I just argued. So zombies can make *zombie* anti-materialist arguments that correspond to the epistemic arguments that people like Chalmers make. If the epistemic arguments are sound, then so are the zombie epistemic arguments. But the zombie epistemic arguments are unsound, since materialism is true in the zombie world. So the epistemic arguments are also unsound.

The key claim of the zombie refutation is that if the epistemic arguments are sound, then so are the zombie epistemic arguments. Recall that the epistemic arguments have the following form:

(E) There is an epistemic gap between physical and phenomenal truths.

(E→O) If there is an epistemic gap between physical and phenomenal truths, then there is an ontological gap, and materialism is false.

(O) ∴ Materialism is false.

The corresponding zombie epistemic arguments have the following form:

(E*) There is an epistemic gap between physical and schmenomenal truths.

(E*→O) If there is an epistemic gap between physical and schmenomenal truths, then there is an ontological gap, and materialism is false.

(O) ∴ Materialism is false.

Both arguments are valid. If (E) and the correspondence thesis are true, then (E*) is true. I discussed the correspondence thesis at length in chapter 3. So here I will focus on (E→O) and (E*→O).

The intuitive challenge for anti-materialists is this. They think that (E→O) is true in the actual world. If they think that (E*→O) is false in the zombie world, then they must say what the relevant difference is between the epistemic gap faced by us and the epistemic gap faced by zombies that would make (E→O) true in the actual world but (E*→O) false in the zombie world. Of course, if zombies' schmenomenal beliefs were false, then there would be an easy answer: only truth-involving epistemic gaps entail ontological gaps. But if the truth thesis is true, then this is not an option for anti-materialists.

Here is a more technical way of spelling out the challenge to anti-materialists. (E→O) claims that one specific epistemic gap, the epistemic gap between physical and phenomenal truths, entails an ontological gap. But anti-materialists' motivation for (E→O) comes from the motivation for a more general principle connecting epistemic gaps to ontological gaps. The idea is that *any* truth-involving epistemic gap entails an ontological gap. The epistemic gap between physical and schmenomenal truths is one such epistemic gap. So if this argument for (E→O) is sound, then a general principle connecting epistemic gaps to ontological gaps is true, and so (E*→O) is true in the zombie world.

It might help to work through an example of the sort of general principle that defenders of might (E→O) appeal to. One such principle, defended by Chalmers (2012), is *fundamental scrutability*: knowing all the fundamental truths (together with some indexical truths) puts you in a position to know all the truths.[9] So if there is a truth that knowing all the physical truths does not put you in a position to know, then fundamental scrutability entails that there are fundamental truths

[9]Balog (1999, p. 508) calls this 'the A Priori Entailment Thesis'.

apart from the physical truths. In other words: fundamental scrutability entails that if there is an epistemic gap between the physical truths and any truth, then materialism is false. Applying this principle to the zombie world, we get (E*→O).

In principle, proponents of the epistemic arguments could claim that fundamental scrutability is true in the actual world even if it is false in other possible worlds. But such a view would be unmotivated, and in practice I know of none who take this view. For example, Chalmers (2012, §2.6) endorses what he calls *generalized* scrutability theses, which are not restricted to the actual world. Those who accept a scrutability thesis without accepting the corresponding generalized scrutability thesis can be challenged to explain why we should think that it is true in the actual world if it is false in other possible worlds. It is not obvious that this challenge can be met.

To sum up: I defended (E*) in chapter 3. And I just argued that if (E→O) is true in the actual world, then (E*→O) is true in the zombie world. So if the standard anti-materialist argument are sound and (E→O) is true in the actual world, then (E*) and (E*→O) are true in the zombie world, which means that the zombie epistemic arguments are sound. Since the zombie epistemic arguments are *not* sound (because materialism is true in the zombie world), the epistemic arguments are not sound either.

Bibliography

Alter, T. (2013). Social Externalism and the Knowledge Argument. *Mind*, *122*(486), 481–496

Alter, T. (2015). The Structure and Dynamics Argument against Materialism. *Noûs*, *50*(4)

Alter, T., & Nagasawa, Y. (Eds.). (2015). *Consciousness in the Physical World: Perspectives on Russellian Monism*. Philosophy of Mind. Oxford University Press.

Audi, P. (2012). Grounding: Toward a Theory of the In-Virtue-Of Relation. *Journal of Philosophy*, *109*(12), 685–711.

Ball, D. (2009). There Are No Phenomenal Concepts. *Mind*, *118*(472), 935–962.

Balog, K. (1999). Conceivability, Possibility, and the Mind-Body Problem. *The Philosophical Review*, *108*(4)

Balog, K. (2012a). Acquaintance and the Mind-Body Problem. In S. Gozzano & C. Hill (Eds.), *New Perspectives on Type Identity: The Mental and the Physical* (pp. 16–42). Cambridge University Press.

Balog, K. (2012b). In Defense of the Phenomenal Concept Strategy. *Philosophy and Phenomenological Research*, *84*(1), 1–23.

Bliss, R., & Trogdon, K. (2016). Metaphysical Grounding. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab,

Stanford University. Retrieved September 28, 2018, from https://plato.stanford.edu/archives/win2016/entries/grounding/

Block, N. (1978). Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science*, *9*, 261–325.

Block, N. (1996). What is Functionalism? In D. M. Borchert (Ed.), *The Encyclopedia of Philosophy Supplement*. MacMillan. Retrieved from http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/functionalism.html

Block, N. (2006). Max Black's Objection to Mind-Body Identity. *Oxford Studies in Metaphysics*, *3*, 3–78.

Block, N., & Stalnaker, R. (1999). Conceptual Analysis, Dualism, and the Explanatory Gap. *The Philosophical Review*, *108*(1)

Boghossian, P. A. (1997). Analyticity. In B. Hale & C. Wright (Eds.), *A Companion to the Philosophy of Language* (pp. 331–368). Blackwell.

Bourget, D., & Chalmers, D. (2014). What Do Philosophers Believe? *Philosophical Studies*, *170*(3), 465–500.

Boutel, A. (2013). How to Be a Type-C Physicalist. *Philosophical Studies*, *164*(2), 301–320.

Broad, C. D. (1925). *The Mind and Its Place in Nature*. Routledge.

Bruntrup, G., & Jaskolla, L. (Eds.). (2016). *Panpsychism: Contemporary Perspectives*. Oxford University Press.

Burge, T. (1979). Individualism and the Mental. *Midwest Studies In Philosophy*, *4*(1), 73–121

Byrne, A. (1999). Cosmic Hermeneutics. *Noûs*, *33*, 347–383

Carruthers, P., & Veillet, B. (2007). The Phenomenal Concept Strategy. *Journal of Consciousness Studies*, *14*(9-10), 212–236.

Chalmers, D. (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, *2*(3), 200–19.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, D. (2002). Does Conceivability Entail Possibility? In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and Possibility* (pp. 145–200). Oxford University Press.

Chalmers, D. (2003a). Consciousness and its Place in Nature. In S. P. Stich & T. A. Warfield (Eds.), *Blackwell Guide to the Philosophy of Mind* (pp. 102–142). Blackwell.

Chalmers, D. (2003b). The Content and Epistemology of Phenomenal Belief. In Q. Smith & A. Jokic (Eds.), *Consciousness: New Philosophical Perspectives* (pp. 220–72). Oxford University Press.

Chalmers, D. (2004). Imagination, Indexicality, and Intensions. *Philosophy and Phenomenological Research*, *68*(1), 182–190

Chalmers, D. (2007). Phenomenal Concepts and the Explanatory Gap. In T. Alter & S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Chalmers, D. (2010a). *The Character of Consciousness*. Oxford University Press.

Chalmers, D. (2010b). The Two-Dimensional Argument Against Materialism. In *The Character of Consciousness* (pp. 141–192). Oxford University Press.

Chalmers, D. (2012). *Constructing the World*. Oxford University Press.

Chalmers, D. (2014). Strong necessities and the mind–body problem: A reply. *Philosophical Studies*, *167*(3), 785–800

Chalmers, D. (2015). Panpsychism and Panprotopsychism. In *Consciousness in the Physical World: Perspectives on Russellian Monism* (pp. 246–276). Philosophy of Mind. Oxford, New York: Oxford University Press.

Chalmers, D. (2016). The Combination Problem for Panpsychism. In G. Brüntrup & L. Jaskolla (Eds.), *Panpsychism*. Oxford University Press.

Chalmers, D., & Jackson, F. (2001). Conceptual Analysis and Reductive Explanation. *Philosophical Review*, *110*(3), 315–61.

Dennett, D. C. (1991). *Consciousness Explained*. Penguin Books.

Descartes, R. (1998). *Discourse on Method* (D. A. Cress, Trans.). Indianapolis, Ind: Hackett.

Deutsch, M. E. (2015). *The Myth of the Intuitive*. The MIT Press.

Diaz-Leon, E. (2010). Can Phenomenal Concepts Explain The Epistemic Gap? *Mind*, *119*(476), 933–951.

Doggett, T., & Stoljar, D. (2010). Does Nagel's Footnote Eleven Solve the Mind-Body Problem? *Philosophical Issues*, *20*(1), 125–143.

Fallon, F. (2016). Integrated Information Theory of Consciousness. In *Internet Encyclopedia of Philosophy*. Retrieved from http://www.iep.utm.edu/int-info/

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press.

Frankish, K. (2016). Illusionism as a Theory of Consciousness. *Journal of Consciousness Studies*, *23*(11-12), 11–39.

Gettier, E. (1963). Is Justified True Belief Knowledge? *Analysis*, *23*(6), 121–123.

Goff, P. (2017). *Consciousness and Fundamental Reality*. Philosophy of Mind. Oxford, New York: Oxford University Press.

Hill, C. S. (1997). Imaginability, Conceivability, Possibility, and the Mind-Body Problem. *Philosophical Studies*, *87*(1), 61–85.

Hill, C. S., & Mclaughlin, B. P. (1999). There are Fewer Things in Reality Than are Dreamt of in Chalmers's Philosophy. *Philosophy and Phenomenological Research*, *59*(2), 445–454

Hume, D. (2011a). *A Treatise of Human Nature* (D. F. Norton & M. J. Norton, Eds.). Clarendon Press.

Hume, D. (2011b). *An Enquiry Concerning Human Understanding*. Hackett.

Ichikawa, J. J., & Steup, M. (2018). The Analysis of Knowledge. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018). Metaphysics Research Lab, Stanford University. Retrieved October 1, 2018, from https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/

Jackson, F. (1982). Epiphenomenal Qualia. *The Philosophical Quarterly*, *32*(127), 127–136

Jackson, F. (1986). What Mary Didn't Know. *The Journal of Philosophy*, *83*(5), 291–295

Jackson, F. (1998). *From Metaphysics to Ethics*. Oxford University Press.

Jackson, F. (2003). Mind and Illusion. *Royal Institute of Philosophy Supplements*, *53*, 251–271

Kelly, T. (2014). Evidence. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2014). Retrieved November 3, 2016, from http://plato.stanford.edu/archives/fall2014/entries/evidence/

Kriegel, U. (2003). Is Intentionality Dependent Upon Consciousness? *Philosophical Studies*, *116*(3), 271–307.

Levin, J. (2007). What is a Phenomenal Concept? In T. Alter & S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Levin, J. (2018). Functionalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Metaphysics Research Lab, Stanford University. Retrieved September 28, 2018, from https://plato.stanford.edu/archives/fall2018/entries/functionalism/

Levine, J. (1998). Conceivability and the Metaphysics of Mind. *Noûs*, *32*(4), 449–480.

Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford University Press.

Levine, J. (2007). Phenomenal Concepts and the Materialist Constraint. In T. Alter & S. Walter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.

Levine, J. (2014). Modality, semantics, and consciousness. *Philosophical Studies*, *167*(3), 775–784

Lewis, D. (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, *50*(3), 249–258.

Lewis, D. (1979). Attitudes de Dicto and de Se. *Philosophical Review*, *88*(4), 513–543.

Lewis, D. (1990). What Experience Teaches. In W. G. Lycan (Ed.), *Mind and Cognition* (pp. 29–57). Blackwell.

Lewis, D. (1994). Reduction of Mind. In S. Guttenplan (Ed.), *Companion to the Philosophy of Mind* (pp. 412–431). Blackwell.

Lewis, D. K. (1966). An Argument for the Identity Theory. *Journal of Philosophy*, *63*(1), 17–25.

Loar, B. (1990). Phenomenal States. *Philosophical Perspectives*, *4*, 81–108

McGinn, C. (1989a). Can We Solve the Mind–Body Problem? *Mind*, *98*(391), 349–366

McGinn, C. (1989b). Consciousness and Content. In *Proceedings of the British Academy, Volume 74: 1988* (pp. 225–245).

Nagel, T. (1974). What is It Like to Be a Bat? *Philosophical Review*, *83*(4), 435–50.

Nida-Rumelin, M. (1995). What Mary Couldn't Know: Belief About Phenomenal States. In T. Metzinger (Ed.), *Conscious Experience* (pp. 219–41). Ferdinand Schoningh.

Nida-Rümelin, M. (2015). Qualia: The Knowledge Argument. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2015). Metaphysics Research Lab, Stanford University. Retrieved September 28, 2018, from https://plato.stanford.edu/archives/sum2015/entries/qualia-knowledge/

Papineau, D. (1993). Physicalism, Consciousness, and the Antipathetic Fallacy. *Australasian Journal of Philosophy*, *71*(2), 169–83.

Papineau, D. (1998). Mind the Gap. *Philosophical Perspectives*, *12*(S12), 373–89.

Papineau, D. (2002). *Thinking about consciousness*. Oxford University Press.

Papineau, D. (2007). Phenomenal and Perceptual Concepts. In S. Walter & T. Alter (Eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (pp. 111–144). Oxford University Press.

Papineau, D. (2011). What Exactly is the Explanatory Gap? *Philosophia*, *39*(1), 5–19

Pereboom, D. (2011). *Consciousness and the Prospects of Physicalism*. Oxford University Press.

Pereboom, D. (2014). Russellian Monism and Absolutely Intrinsic Properties. In U. Kriegel (Ed.), *Current Controversies in Philosophy of Mind* (pp. 40–69). Routledge.

Pereboom, D. (2015). Consciousness, Physicalism, and Absolutely Intrinsic Properties. In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism*. Oxford University Press.

Perry, J. (1979). The Problem of the Essential Indexical. *Noûs*, *13*(1), 3–21.

Perry, J. (2001). *Knowledge, Possibility, and Consciousness*. MIT Press.

Prior, A. N. (1960). The Autonomy of Ethics. *Australasian Journal of Philosophy*, *38*(3), 199–206.

Putnam, H. (1963). Brains and Behavior. In R. J. Butler (Ed.), *Analytical Philosophy: Second Series*. Blackwell.

Rosen, G. (2010). Metaphysical Dependence: Grounding and Reduction. In B. Hale & A. Hoffmann (Eds.), *Modality: Metaphysics, Logic, and Epistemology* (pp. 109–36). Oxford University Press.

Russell, G., & Restall, G. (2010). Barriers to Implication. In C. Pigden (Ed.), *Hume on Is and Ought*. Palgrave MacMillan.

Schaffer, J. (2017). The Ground Between the Gaps. *Philosophers' Imprint*, *17*(11).

Searle, J. R. (1992). *The Rediscovery of the Mind*. MIT Press.

Shoemaker, S. (1975). Functionalism and qualia. *Philosophical Studies*, *27*(5), 291–315

Shoemaker, S. (1999). On David Chalmers's the Conscious Mind. *Philosophy and Phenomenological Research*, *59*(2), 439–444

Stoljar, D. (2001). Two Conceptions of the Physical. *Philosophy and Phenomenological Research*, *62*(2), 253–281

Stoljar, D. (2005). Physicalism and Phenomenal Concepts. *Mind and Language*, *20*(2), 296–302.

Stoljar, D. (2006). *Ignorance and imagination: On the epistemic origin of the problem of consciousness*. Oxford University Press.

Stoljar, D. (2010). *Physicalism*. Routledge.

Stoljar, D. (2014). Four Kinds of Russellian Monism. In U. Kriegel (Ed.), *Current Controversies in Philosophy of Mind* (pp. 17–39). Routledge.

Stoljar, D. (2015). Russellian Monism or Nagelian Monism? In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism*. Oxford University Press.

Stoljar, D. (2017). Physicalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017). Metaphysics Research Lab, Stanford University. Retrieved September 28, 2018, from https://plato.stanford.edu/archives/win2017/entries/physicalism/

Sundström, P. (2008). Is the Mystery an Illusion? Papineau on the Problem of Consciousness. *Synthese*, *163*(2), 133–143.

Tononi, G. (2007). The Information Integration Theory of Consciousness. In M. Velmans & S. Schneider (Eds.), *The Blackwell Companion to Consciousness*. Blackwell.

Tye, M. (1999). Phenomenal Consciousness: The Explanatory Gap as a Cognitive Illusion. *Mind*, *108*(432), 705–725.

Tye, M. (2000). *Consciousness, Color, and Content*. MIT Press.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.