

A Response to Chalmers' Master Argument¹

Mattan Griffel

New York University

Introduction

Anti-physicalists sometimes argue from certain epistemic gaps to the existence of an ontological gap in nature. The most promising physicalist response has become known as the “phenomenal concept strategy”. Chalmers’ (2007) “master argument” claims to show that *no* version of the phenomenal concept strategy can succeed. In this paper I argue that Chalmers’ master argument mischaracterizes the phenomenal concept strategy. I conclude by proposing an alternative characterization of the phenomenal concept strategy, one that is immune to Chalmers’ objections.

1. The epistemic gaps

Let P be the complete physical truth about our world: a really long sentence describing all the physical properties of every physical thing in our universe as well as the physical laws by which they are governed. Let Q be an arbitrary truth about

¹ I would like to thank my advisor Ned Block for his generosity and infinite patience, as well as Kati Balog, David Chalmers, Diane Klein, Harrison Mahaffey, Farid Masrou, James Partridge, Jim Pryor, Ted Sider and Sharon Street for comments and helpful discussion of this paper.

consciousness²: for example, the fact that someone is conscious, or that I am having a phenomenal experience of a certain shade of red.

Most puzzles about consciousness start with the observation that there appear to be gaps between *P* and *Q*. Let's consider three thought-experiments:

TE 1³: Mary has spent her entire life inside a black-and-white room. From inside her room she has somehow come to know all of the physical facts about the world, perhaps by reading books or by looking out at the world using a black-and-white television screen. Mary knows *P*. But what happens when she's released from her room? We can try to imagine her reaction as she sees colors for the first time. "So *this* is what it's like to see red!" she might say. Having a phenomenal experience of redness for the first time... does this mean Mary has learned something new? Suppose *Q* is a fact about what it is like to see red. Does Mary now know *Q*? If not, there was a gap between her knowledge of *P* and her knowledge of *Q*.

TE 2: A zombie is a hypothetical creature that is physically identical to a conscious being but is not itself conscious. Can you conceive of a zombie? There certainly doesn't seem to be any sort of contradiction. Consider a zombie world that is physically identical to ours, but in which *no* creature is conscious. Suppose *Q* is the truth that someone is conscious. Zombie world is a world in which *P* is true but *Q* is false. Can you conceive of zombie world? If so, there is a gap between your conceiving of *P* and conceiving of *Q*.

² By "consciousness" I mean subjective or phenomenal experience. A thing is "conscious" if there is *something it is like* be that thing. This definition of consciousness was proposed by Nagel (1974).

³ TE 1 is a variation of Frank Jackson's case of Mary in the black-and-white-room.

TE 3: The natural sciences – biology, chemistry, physics – are all the ways in which we try to understand the world around us. Imagine being given an object you’ve never seen before: it looks like a cross between a stool and a banana, plus it can fly. How would you explain it physically? You can use every possible scientific tool available. You could investigate its structure: spatiotemporal features like weight, density, composition. You could investigate its function: causal roles like how it interacts with its environment, what it does, how it is able to fly. Now imagine being given Q : the way a particular cup of coffee smells. It doesn’t seem to have a structure. It doesn’t seem to have a function. Could you explain Q to me in physical terms? If not, there is a gap between P and explainability of Q .

Each of these thought experiments represents a particular kind of *epistemic gap*. They’re called epistemic gaps because each of them denies an *epistemic relation* between P and Q : a relation involving what we can *know*, what we can *conceive*, or what we can *explain*. They can be summarized much more clearly (though much less vividly) like this:

The knowledge gap: Q cannot be deduced from P

The conceptual gap: P can be conceived without Q

The explanatory gap: P cannot explain Q

2. Inference to an ontological gap

From these three epistemic gaps, some people infer that there is a corresponding gap in the fundamental nature, or ontology, of our world. Let’s call

these people “anti-physicalists”. Anti-physicalists tend to disagree with another group of people: “physicalists”. Physicalists think that the world is purely physical. Anti-physicalists do not. Anti-physicalists infer the existence of an ontological gap in nature between the physical and the phenomenal, that is, they infer that the physical and the phenomenal are two fundamentally different kinds of things. The specific inferential claim an anti-physicalist will make depends on which epistemic gap she starts out from. This inference to an ontological gap is either direct or indirect. *Direct inference* happens on the basis of the explanatory gap: if physical facts can’t explain consciousness, then consciousness can’t be a physical fact. An *indirect inference* happens on the basis of the other two epistemic gaps. For example, if $P \& \sim Q$ is conceivable, anti-physicalists often infer that $P \& \sim Q$ is metaphysically possible, and therefore physicalism is false.

Chalmers (1996) argues that this kind of an inference is justified on *a priori* grounds because reason itself entitles us to make it. The idea is that, by default, if S is conceivable then we should believe that S is possible. We *shouldn’t* understand Chalmers to be saying this inference will always be right. In many cases, no doubt, this kind of inference leads us to the wrong conclusion. But if we do it carefully, taking all the necessary precautions to avoid making the inference when we shouldn’t, then Chalmers thinks we can be pretty certain that we’ll be right (certain enough to qualify as being justified). The burden of proof, then, lies on the physicalist to provide some sort of indication that the inference to an ontological gap is a bad one. And if she can’t, then why *not* infer?

Arguments like those outlined above that rely on inference from an epistemic gap to an ontological gap, are what Chalmers (2002a) calls *epistemic arguments* against physicalism. Roughly speaking, they work like this:

- (1) There is an epistemic gap between physical and phenomenal truths.
- (2) If there is an epistemic gap between physical and phenomenal truths, then there is an ontological gap, and physicalism is false.

-
- (3) Physicalism is false.

3. The type-B physicalist response

Type-B physicalists accept the epistemic gaps but deny an ontological gap. They hold that Mary lacks knowledge but not knowledge of any non-physical facts, that zombies are conceivable but not metaphysically possible, and that although there will never be a satisfying explanation of consciousness in terms of physical processes, consciousness is a result of physical processes nonetheless. This view may strike some as ad hoc. If consciousness *is* a result of physical processes, why think that the process by which it occurs is epistemically inaccessible? There doesn't seem to be any satisfying way to ground the existence of the epistemic gaps in a universe that is entirely physical.

Type-B physicalists face the difficult task of denying the inference from the epistemic gaps to an ontological gap. One way to do this is to deny that our modal intuitions – intuitions about what scenarios are possible or necessary – are ever

justified. This would certainly help type-B physicalists deal with the problems posed by the epistemic gaps, but it would also require the type-B physicalist to defend something like a very strong form of *modal skepticism*, the view that we are not justified in making certain modal inferences to metaphysical possibility or necessity.⁴ Even among modal skeptics, few are willing to argue that our modal intuitions are *never* justified. Surely, we are at least sometimes justified in making modal inferences in cases of everyday life, or in science, or even in philosophy. We often use our modal intuitions as the foundation for beliefs about the possible consequences of our actions. For example, when I look out the window and see dark clouds, I form the belief that it might rain, and so I grab my umbrella before going outside. Would the type-B physicalist really want to say that all of our beliefs about the future formed on the basis of our modal intuitions are unjustified? This strikes me as a bad option.

What the type-B physicalist needs to do is discriminate between those cases in which our modal intuitions are justified and those cases in which they are not. Hopefully, she can then show why our modal intuition in the case of the inference from the epistemic gaps to an ontological gap is not justified. This is exactly what the type-B physicalist attempts to do in the phenomenal concept strategy.

4. The phenomenal concept strategy

From here on I'll be focusing on the phenomenal concept strategy. The discussion will consist of a brief overview of its origins, and a review of some of the

⁴ For more on the issues of modal skepticism see Van Inwagen's "Modal Epistemology" (1997).

more prominent accounts. Though for the most part I'll be talking about the strategy in broad strokes. The idea is to get a good feel for what exactly it tries to do and how it tries to do it. That way when we get to Chalmers' attack on the phenomenal concept strategy, we'll know what to expect and how to deal with it.

The basic idea behind the phenomenal concept strategy is to locate the source of the epistemic gaps in the *epistemic* relationship between our physical and phenomenal concepts rather than the *metaphysical* relationship between physical and phenomenal properties. Proponents of the phenomenal concept strategy provide various accounts of the nature of our phenomenal concepts, and then they suggest that these accounts *predict* the existence of the epistemic gaps between physical properties conceived under physical concepts and phenomenal properties conceived under phenomenal concepts. Most importantly, they claim that these accounts are compatible with physicalism. If all of these steps are done correctly, the phenomenal concept strategy should show how the epistemic gaps are compatible with physicalism.

The phenomenal concept strategy developed out of a paper by Brian Loar called "Phenomenal States" (1990/97) in which he suggests that phenomenal concepts are unlike other concepts in the following two ways: (i) they *refer directly* to phenomenal states (their *referents*), and (ii) their *mode of presentation*, or way in which they present their referent, somehow involves the phenomenal experience itself. Subsequent accounts of phenomenal concepts tend to develop either of these

two ideas, so it makes sense to classify them into two categories: *direct-reference accounts* and *special modes of presentation accounts*.⁵

4.1. Direct-reference accounts

Direct-reference accounts focus on the directness of the relationship between phenomenal concepts and their referents, phenomenal states:

On the *causal-recognitional account*, Tye (2003) suggests that phenomenal concepts are special recognitional concepts that refer directly. They have no mode of presentation. Instead, they are tied to their referents via a causal connection.

On the *demonstrative account* phenomenal concepts are a sort of demonstrative. Perry (2001) suggests that phenomenal concepts are demonstratives where the demonstrative is guided to its referent via a perceptual state. Levin (2007) suggests that phenomenal concepts are type demonstratives with no mode of presentation at all.

On the *information-theoretic account*, Aydede and Güzeldere (2005) suggest an information-theoretic analysis of the relation between phenomenal concepts and sensory concepts. They claim we are cognitively designed to acquire sensory concepts directly from our experiences. (e.g. concepts of colors, sounds, shapes are triggered by corresponding experiences). It is suggested that these sensory concepts double as phenomenal concepts when we use the same cognitive structures in introspection.

⁵ For the classifications in this section, I am much indebted to the work done by Balog (2009a)

4.2. Special modes of presentation accounts

Special modes of presentation accounts try to capture the special intimacy between phenomenal concepts and phenomenal states by suggesting that the mode of presentation of a phenomenal concept involves the phenomenal experience itself that the phenomenal concept refers to:

On the *recognitional account*, Carruthers (2004) suggests that phenomenal concepts are recognitional concepts without descriptive modes of presentation and posits the existence of higher-order experiences of experiences that guide our phenomenal concepts to their referents.

On the *constitutional account*, phenomenal concepts are constituted by the phenomenal experiences they refer to. For most concepts, it doesn't matter what constitutes a particular token of a concept, so long as the right kinds of causal or informational relations hold between it and the rest of the world. For example, it doesn't matter what neural configurations constitute a particular token of the concept *bird* as long as the right kinds of causal or informational relations between *bird* and birds hold. However, for phenomenal concepts constitution *does* matter for reference: both in terms of how reference is determined and how the concepts present their referents. Every token of a phenomenal concept applied to a current phenomenal experience is constituted by that phenomenal experience itself. Versions of the constitutional account of phenomenal concepts have been proposed by, Papineau (2002, 2007), Block (2006), and Balog (2006).

On the *distinct conceptual roles account*, first proposed by Nagel (1974) and defended by Hill (1997) Hill and McLaughlin (1999), phenomenal concepts and

physical concepts are employed by independently-operating cognitive faculties, and therefore play very different conceptual roles. Our ability to splice together situations conceived under separate cognitive faculties into one seemingly coherent situation explains why we can conceive of physical states without their accompanying phenomenal states and vice versa. As a result, we should expect phenomenal concepts and physical concepts to come apart epistemically, even if they do in fact share a referent.

Various other versions have been proposed by Sturgeon (1994), Ismael (1999), Levine (2001), O'Dea (2002), and Perry (2001).

It is argued that under both direct-reference accounts and special modes of presentation accounts, phenomenal concepts refer to phenomenal states in an intimate and essential manner without revealing anything about their referents' nature that lends itself to physical or functional analysis, either because they have no mode of presentation, or because their mode of presentation is a particular phenomenal experience itself. As a result, the accounts above are entirely compatible with physicalism. Furthermore, it is argued that this intimate relation between phenomenal concepts and phenomenal states explains the presence of the relevant epistemic gaps.

If the phenomenal concept strategy holds up, it will deal a strong blow to anti-physicalism. For one thing, its mere possibility provides a relevant alternative explanation for the epistemic gaps, undercutting some of the justification for the anti-physicalists' inference to an ontological gap. Furthermore, if the phenomenal concept strategy turns out to be true (it is verifiable, after all), then the presence of

the relevant epistemic gaps will be entirely explained. This would completely undercut the anti-physicalists' inference to an ontological gap. Although the anti-physicalist could still believe in the existence of an ontological gap – nothing about the truth of the phenomenal concept strategy rules out the existence of an ontological gap – it seems as if the anti-physicalist would be left with almost no way to argue for her position. The truth of the phenomenal concept strategy would turn the tide decidedly in the physicalists' favor.

An interesting side note, since the accounts of phenomenal concepts themselves take no stance on the ontological nature of phenomenal properties, they are also compatible with anti-physicalism. Chalmers (2003) himself has proposed a version of the constitutional account of phenomenal concepts (although without any associated ambition to support type-B physicalism). His account ends up looking very similar to the physicalist constitutional accounts with the exception that phenomenal concepts are constituted by non-physical states. In response, Balog (2009a) makes a great point: given that Chalmers has endorsed a constitutional account of phenomenal concepts, the phenomenal concept strategy says that there are now two parallel anti-physicalist explanations of the epistemic gaps. The anti-physicalist says that the epistemic gaps exist because of a corresponding ontological gap. But this explanation ends up being redundant, since there is an explanation for the existence of these very same epistemic gaps in terms of the special nature of phenomenal concepts. I think it could be argued that Chalmers' adoption of a constitutional account of phenomenal concepts undermines his reasons for being an anti-physicalist in the first place.

5. Chalmers' master argument

Chalmers' master argument (2007) claims to show that any attempt to provide a physical explanation for the existence of the relevant epistemic gaps between *P* and *Q* will end up creating a second-order explanatory gap. The idea is that even if the phenomenal concept strategy manages to explain the relevant epistemic gaps – which include the original (i.e. “first-order”) explanatory gap – the type-B physicalist will now have to provide some sort of explanation for the second-order explanatory gap. If Chalmers is right, then the phenomenal concept strategy cannot, by itself, successfully ground a type-B physicalist view consistent with the epistemic gaps.

The master argument can be divided into two parts: first, Chalmers represents the general structure of the phenomenal concept strategy, and second, he argues that no account of phenomenal concepts can satisfy that structure.

Chalmers claims the general structure of the phenomenal concept strategy can be represented in the following way. Proponents put forward a thesis *C* claiming that human beings have some sort of “key” psychological features. These “key” features are whatever features the phenomenal concept strategy needs to attribute in order to explain the epistemic gaps. They then argue that: (1) *C* is true; (2) *C* explains our epistemic situation with regard to consciousness; and (3) *C* is physically explicable. If proponents of the phenomenal concept strategy can do all this, Chalmers thinks they will have undercut the inference from the epistemic gaps to the ontological gap.

Chalmers' master argument combines the original arguments for the conceptual gap and the explanatory gap. The hard part of his argument consists in showing that zombies do not share our epistemic situation with regard to consciousness. Once he's done this, Chalmers poses the following dilemma: can one conceive of a zombie world in which C is false? If so, then $P \& \sim C$ is conceivable, and P cannot explain C . If not, then $P \& \sim C$ is inconceivable, and C cannot explain our epistemic situation with regard to consciousness. Either C is not physically explicable or C cannot explain our epistemic situation with regard to consciousness, and so Chalmers concludes that the phenomenal concept strategy fails.

Before we delve into the argument itself let's go over the terminology and some assumptions Chalmers makes:

First, Chalmers takes it that our "epistemic situation" includes the *truth-values* of our beliefs (i.e. whether they are true or false) as well as the *epistemic status* of our beliefs (i.e. whether they are justified or unjustified, cognitively significant or insignificant⁶). On this account, two beings will share an epistemic situation when they have corresponding beliefs with corresponding truth-values and epistemic status. It's important to note that Chalmers claims to assume an intuitive notion of 'correspondence' between beliefs: one that does not require corresponding beliefs to have the same *content*. Whether such a view is *actually* intuitive is a point to which I'll return.

⁶ As far as I can tell, Chalmers intends the term "cognitively significant" to apply to beliefs not inferable from beliefs about physical facts. In Jackson's case of Mary in the black-and-white room, Chalmers supposes that Mary gains cognitively significant knowledge of what it is like to see red while her zombie twin, Zombie-Mary does not.

Second, Chalmers defines “conceivability” as follows:

Conceivability: For any statement S , S is conceivable iff S cannot be ruled out through *a priori* reasoning.⁷

Third, Chalmers assumes a connection between conceivability and explanation. More precisely, he thinks that there is a connection between conceivability and a certain sort of reductive explanation: the kind that makes it transparent why some high-level truth obtains given that some low-level truths obtain. According to Chalmers:

CE Principle: For any high-level truth, ϕ , and any low-level truth, ψ , if $\psi \& \sim \phi$ is conceivable then ψ cannot explain ϕ .

“ ψ cannot explain ϕ ” means there is no explanation that makes it transparent why ϕ obtains given that ψ obtains. This makes sense, given that our notion of conceivability requires that we be unable to rule a thing out *a priori*. If we can’t rule out $\psi \& \sim \phi$ *a priori*, then there’s no *transparent* explanation for why ϕ should be true given that ψ is true.

⁷ This notion of conceivability corresponds to what Chalmers calls *negative conceivability*. Chalmers sometimes appeals to a stronger notion of conceivability, *positive conceivability*, according to which S is conceivable iff one is able to form a ‘clear and distinct’ conception of a situation in which S is true. For more on the issue, see Chalmers (2002b).

Finally, Chalmers assumes that zombies come with the standard set of intentional mental states; roughly speaking, zombies can have mental states *of* or *about* things. This means that zombies have (or rather, one can conceive of them as having) the ability to form beliefs and possess concepts. I want to take a little time to consider this important point.

One might object to Chalmers' assumption that a non-conscious being could possess intentional mental states by claiming that consciousness is necessary for intentionality. Searle (1992) argues that non-conscious mental states must be 'potentially conscious'. In this case, although zombies' physical brain states may resemble ours, they don't have any potentially conscious mental states, so they lack intentionality. If zombies lack intentionality, then they can't form beliefs or possess concepts.

In opposition to Searle, I think there are three reasons to believe that zombies can have intentional mental states. First, all of the usual principles of psychological explanation support the attribution of beliefs to zombies. Presuming that physics is causally closed, my zombie-counterpart and I are behaviorally identical. If we both say "I am hungry" and then both proceed to open the fridge and make a sandwich, our behavior is best explained by the fact that we believe we are hungry, believe that a sandwich will sate our hunger, and believe that the fridge contains the ingredients necessary for making a sandwich. Explaining my zombie-counterpart's behavior without attributing to it beliefs would be quite difficult.

Second, zombies possess physical brain states identical to ours with the same kinds of causal relations linking them to the world, which, as Balog (1999)

notes, would count zombies as having intentionality on all the current theories of meaning. On a Davidsonian interpretationist account, zombies will have intentionality because they are just as interpretable as conscious beings. The same goes for such theories as the informational account (e.g. Dretske 1988), the causal-historical account (e.g. Kripke 1972), the counterfactual account (e.g. Fodor 1990), the teleosemantic account (e.g. Millikan 1989, Papineau 1993), etc. The only account on which zombies do not count as having intentionality is the account on which consciousness is required.

Third, there doesn't seem to be any reason to think that beliefs are necessarily phenomenal. It is certainly possible for a conscious being to attend to a belief about x , and thereby have a phenomenal experience of oneself as *believing* x . But if attention were necessary for the possession of a belief, one would cease believing something the second one stopped attending to the belief. For example, if I ask you whether you believe that the moon is made of cheese and you say no, and then I ask you again ten minutes later and you say no, then we shouldn't say that for those ten minutes you stopped believing the moon was not made of cheese. In any case, both Chalmers and the type-B physicalist hold that zombies can have intentional mental states, so let us set the issue aside.

Let's now evaluate the master argument.

Chalmers wants to show that zombies do not share our epistemic situation with regard to consciousness. For a given conscious being with a given epistemic situation, let E be a sentence asserting the existence of a being with that epistemic situation. E will be made true by that conscious being in its original epistemic

situation, and E will also be made true by any being that shares this epistemic situation. Supposing that the conscious being considered in E has at least one true belief about consciousness, that true belief can be represented as Q . For example, the conscious being might have a true belief *someone is conscious* or *I am having the phenomenal experience of a certain shade of red*. But what about her zombie-counterpart in zombie world? It is plausible that a non-conscious being can't have beliefs with exactly the same content as our beliefs about consciousness, but given the fact that our notion of 'correspondence' between beliefs is content-neutral, Chalmers thinks we can nevertheless talk of the zombie's corresponding beliefs. So the conscious being with the true belief Q will have a zombie-counterpart in zombie world with the belief Q . But the zombie-counterpart's belief will be false. For example, if the conscious being has the true belief *someone is conscious* then her zombie-counterpart in zombie world has the corresponding false belief *someone is conscious*. Given that our epistemic situation includes the truth-values of our beliefs, the two will not share an epistemic situation with regard to consciousness. E is true in our world but false in zombie world. If zombie world is conceivable, the argument above has managed to show that $P \& \sim E$ is conceivable.

Given that $P \& \sim E$ is conceivable, the rest of the master argument is quite straightforward. Either $P \& \sim E \& \sim C$ is conceivable or $P \& \sim E \& C$ is inconceivable. This is logically true in virtue of its form. If $P \& \sim E \& \sim C$ is conceivable, then $P \& \sim C$ is conceivable, and applying the CE Principle tells us that P cannot explain C . If $P \& \sim E \& C$ is inconceivable, since it is given that $P \& \sim E$ is conceivable, when we are conceiving of $P \& \sim E$ we must be conceiving of a world in which C is true. In this case,

$C \& \sim E$ is conceivable, and applying the CE Principle tells us that C cannot explain E .

We can summarize the argument above as follows:

- (1) $P \& \sim E$ is conceivable
 - (2) If $P \& \sim E$ is conceivable then $P \& \sim C$ is conceivable or $C \& \sim E$ is conceivable
 - (3) If $P \& \sim C$ is conceivable then P cannot explain C
 - (4) If $C \& \sim E$ is conceivable then C cannot explain E
-
- (5) P cannot explain C or C cannot explain E

6. Balog's defense of the phenomenal concept strategy

Balog (2009b) thinks that $P \& \sim C$ is conceivable and embraces the conclusion that P cannot explain C . She argues that even if the phenomenal concept strategy produces a second-order explanatory gap, this second-order explanatory gap doesn't pose a problem for the physicalist because it can be explained by the phenomenal concept strategy itself. The phenomenal concept strategy, if it is true, predicts an explanatory gap between physical facts and phenomenal facts. So, if we think about C in phenomenal terms then, if C is true, we should *expect* a second-order explanatory gap between the physical facts and C .

I think this is an interesting move. Chalmers charges that the explanatory scheme outlined above is circular. Balog doesn't deny this. She accepts that her argument is circular, but denies that there's anything *viciously* circular about it. Balog is essentially saying that if C is physical then C explains the second-order

explanatory gap between *P* and *C* in a way that is compatible with physicalism. From the type-B physicalist's perspective (start by assuming *C* is physical), Balog's circular argument allows the phenomenal concept strategy to successfully undercut the anti-physicalist inference to an ontological gap. However, from Chalmers' perspective (assuming that the inference to an ontological gap is justified), the second-order explanatory gap between *P* and *C* would entail an ontological gap between *P* and *C*, thereby successfully neutralizing the force of the phenomenal concept strategy. Both Balog and Chalmers are right.

Interestingly, the situation above is a sort of metaphysical stalemate between the anti-physicalist and the physicalist. Each side can unseat the other, if permitted one core assumption. Balog concludes that there's a puzzling symmetry between both positions. This is not a conclusion I'm satisfied with.

7. Attacking Chalmers' master argument

Instead, I propose accepting that *C* cannot explain *E* but arguing that the phenomenal concept strategy will still work. In particular, I intend to show that Chalmers' has mischaracterized the phenomenal concept strategy and, in doing so, constructed a straw-man argument.

Why should *C* explain our epistemic situation with regard to consciousness? In the following passage, Chalmers claims that requiring anything less of the phenomenal concept strategy would undercut its force in supporting type-B physicalism:

Recall that the strategy is intended to resist the antiphysicist's inference from an epistemic gap to an ontological gap by showing how the relevant epistemic gap may exist even if physicalism is true. In the antiphysicist's arguments, the relevant epistemic gap (from which the ontological gap is inferred) is characterized in such a way that truth and knowledge are essential.... If one characterized these gaps in a way that were neutral on the truth of phenomenal beliefs, the 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.** (Chalmers 2007)

In other words, the anti-physicalist arguments for the relevant epistemic gaps pose a problem for type-B physicalists because they are truth-involving: they are epistemic gaps between *P* and *Q* where both are *true* of our world. If we characterize the arguments in such a way that either *P* or *Q* is false, then they won't support the existence of a *truth-involving* epistemic gap. And the anti-physicalist can't infer an *truth-involving* ontological gap from a *truth-neutral* epistemic gap. To undercut the inference to an ontological gap, Chalmers claims that *C* will have to provide an alternative explanation of the truth-involving epistemic gaps.

How Chalmers gets from here to the claim that *C* must explain our epistemic situation with regard to consciousness isn't exactly clear. That is to say, I'm unclear about the relationship between our epistemic situation and the relevant epistemic gaps. Presumably, the epistemic gaps are a part of our epistemic situation in general, perhaps even with regard to consciousness. Even so, one could argue that, by hypothesis, our epistemic situation with regard to consciousness includes beliefs about consciousness that are unrelated to the epistemic gaps, and so should not have to be explained by *C*. For example, my belief that *either I am conscious or I am*

not conscious is a belief “about” consciousness (at least, apparently), and so is part of my epistemic situation in regard to consciousness. But I would imagine that an explanation for the truth and justification of my belief that *either I am conscious or I am not conscious* should include that it is true and justified in virtue of its logical form. Such an explanation would then have to include facts about logical form, tautology, etc. There’s no reason to think that *C* should have to explain things like logical form and tautology in order to explain the presence of the truth-involving epistemic gaps. This supports the conclusion that requiring *C* to explain our epistemic situation with regard to consciousness is too strong.

On top of that, I think that there are reasons to think that even expecting *C* to explain the relevant epistemic gaps is too strong. First, note that the bolded section in Chalmers’ excerpt above reads: “the phenomenal concept strategy needs to show how the relevant truth-involving epistemic gaps are consistent with physicalism.” This is **not** equivalent to saying that the phenomenal concept strategy needs to show how physical facts can explain the relevant truth-involving epistemic gaps. But it **is** equivalent to saying that the phenomenal concept strategy needs to show how, in a purely physical world, the relevant truth-involving epistemic gaps can exist. This second claim says nothing about physical facts explaining the epistemic gaps, which gives us reason to think that requiring *C* to explain the relevant epistemic gaps is too strong.

Second, Chalmers emphasizes that what is important about the relevant epistemic gaps is that they are truth-involving. Consider the truth-involving conceptual gap. It would be tempting but incorrect to express it as ‘ $P \& \sim Q$ ’ is

conceivable,' because this wouldn't account for the fact that P and Q are both true in our world. We might say that ' $P \& \sim Q$ is conceivable' expresses a *truth-neutral* conceptual gap. But we can turn this truth-neutral conceptual gap into the truth-involving conceptual gap by adding in the truth of P and Q . In other words, the proper way to express the truth-involving conceptual gap is: $P \& Q$ and $P \& \sim Q$ is conceivable. If this is the case, then to require C to explain the presence of the truth-involving epistemic gaps is to require C to explain not only $P \& \sim Q$ is conceivable but also P and Q . No type-B physicalist could accept the requirement that C explain Q , because the type-B physicalist holds that nothing physical can explain Q . In fact, it turns out that the only way that C could explain the truth-involving epistemic gaps is if C could explain P as well. But given that P is the entire fundamental physical truth about our world, C could explain P only if $C = P$. This would put the proponent of the phenomenal concept strategy right back to where she started.

What I've shown is that, by requiring the proponent of the phenomenal concept strategy to provide some account that explains E , Chalmers has snuck in the requirement that this account explain Q . It's no wonder then that the proponent of the phenomenal concept strategy finds herself facing a second-order explanatory gap. The second-order explanatory gap is really just the first-order explanatory gap in disguise.

Let us step back and reflect on the dialectic as it stands.

Any argument for the relevant truth-involving epistemic gaps, by Chalmers' own admission, must appeal to the truth of P and of Q . For example, to get to the truth-involving conceptual gap – $P \& Q$ but $P \& \sim Q$ is conceivable – the anti-physicalist

must first show that *P* and *Q* are true. He does this implicitly by building truth into their definitions ('Let *P* be the complete fundamental physical truth...' and 'Let *Q* be any arbitrary truth...').

I think the proponent of the phenomenal concept strategy should have access to *P* and *Q* as well. It is the ontological nature of *Q* that is called into question in the debate between physicalists and anti-physicalists, *not* the truth-value of *Q*. Anyone who accepts the existence of phenomenal consciousness will accept that *Q* is true. So I propose that the proponent of the phenomenal concept strategy be able to use *Q* as well in her explanation for the relevant truth-involving epistemic gaps, so long as it's just the truth-value of *Q* that is playing an explanatory role.

This illustrates the final major flaw in Chalmers' characterization of the phenomenal concept strategy: he hasn't distinguished between the strategy itself and the account, *C*, of phenomenal concepts that it provides. *C* is not the only fact available to the phenomenal concept strategy. At the very least, both *P* and *Q* are available as well.

8. An alternative characterization of the phenomenal concept strategy

In this section, my goal will be to provide a more accurate and charitable representation of the general structure of the phenomenal concept strategy: one that both undercuts the inference from the relevant truth-involving epistemic gaps to an ontological gap and is immune to Chalmers' master argument.

Proponents of the phenomenal concept strategy provide a thesis *C** claiming that phenomenal concepts are unlike other concepts in virtue of some direct or

intimate access to their referents. They then argue the following: (1) C^* is true, (2) C^* explains the *conceptual independence* between the physical and the phenomenal, and (3) C^* is physically explicable.

Again, according to Chalmers, “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.”

I contend that given C^* , P , Q , and the CE Principle, the phenomenal concept strategy can show how all the relevant truth-involving epistemic gaps are consistent with physicalism. And if proponents of the phenomenal concept strategy can do all this, on Chalmers’ account they will have undercut the inference from the epistemic gaps to the ontological gap.

Note that my representation of the general structure of the phenomenal concept strategy shares (1) and (3) with Chalmers’. My first task will be to show why, on my account, (2) provides the most charitable and accurate representation of the role C^* is supposed to be playing in the phenomenal concept strategy.

All versions of the phenomenal concept strategy rest on an account C^* claiming that phenomenal concepts are unlike other concepts in virtue of having some sort of intimate or direct relation to phenomenal states. What results are various accounts of phenomenal concepts that refer to phenomenal states in an intimate and substantial manner without revealing anything about their nature that lends itself to physical or functional analysis.

If C^* is true, the fact that phenomenal concepts reveal nothing about their referents’ nature that lends itself to physical or functional analysis could be used to

defend any number of claims about what sorts of epistemic gaps will obtain between physical and phenomenal concepts. However, I believe it will suffice to show here how C^* can explain what I call “conceptual independence” between physical concepts and phenomenal concepts, which I define as follows:

Conceptual independence: for any concept, ψ , and any concept, ϕ , there is a conceptual independence between ψ and ϕ iff the following three scenarios are conceivable: (i) $\psi \& \phi$, (ii) $\sim \psi \& \phi$, and (iii) $\psi \& \sim \phi$.

I think establishing conceptual independence between physical and phenomenal concepts is the most intuitive way to understand the role that the account of phenomenal concepts is supposed to play in the phenomenal concept strategy. For one, it strikes me as obvious that the primary goal of the account of phenomenal concepts used in the phenomenal concept strategy is to establish a sort of conceptual separation between physical and phenomenal concepts. On top of this, given Chalmers’ definition of conceivability, which requires that one be unable to rule out a scenario *a priori*, the fact that phenomenal concepts reveal nothing physically or functionally analyzable about their referents suggests an inability to *a priori* rule out scenarios involving phenomenal and physical concepts. If phenomenal concepts reveal nothing physical or functional about their referents, then there can be no physical or functional basis on which to say that a scenario involving any two concepts, one physical and one phenomenal, can be ruled out *a priori*.

I hope the reader will be satisfied with my notion of conceptual independence between physical and phenomenal concepts, which I take to be a very modest conclusion to draw from the phenomenal concept strategy. Note that I have assumed nothing about the epistemic relations of explanation or knowledge between physical and phenomenal concepts.

My second task will be to prove that the phenomenal concept strategy explains the presence of the three truth-involving epistemic gaps given the following four assumptions: P is true, Q is true, C^* is true, and the CE Principle holds.

(i) The conceptual gap: P can be conceived without Q

I will represent the truth-involving conceptual gap as follows: $P \& Q$ and $P \& \sim Q$ is conceivable. Given P , Q , and C^* , the phenomenal concept strategy can show how the truth-involving conceptual gap is consistent with physicalism. If C^* is true, then P and Q are conceptually independent, and so the following three scenarios are conceivable: $P \& Q$, $P \& \sim Q$, $\sim P \& Q$. The one that matters for the purpose of explaining the conceptual gap is that $P \& \sim Q$ is conceivable. Since it is given that P is true and Q is true, we can conclude that $P \& Q$ and $P \& \sim Q$ is conceivable (i.e. there will be a conceptual gap between P and Q)

(ii) The explanatory gap: P cannot explain Q

I will represent the truth-involving explanatory gap as follows: $P \& Q$ but P cannot explain Q . Given P , Q , C^* , and the CE Principle, the phenomenal concept strategy can show how the explanatory gap is consistent with physicalism. From C^*

we deduced that $P \& \sim Q$ is conceivable. By applying Chalmers' CE Principle, we can conclude that P cannot explain Q . So given P, Q, C^* , and the CE Principle, the phenomenal concept strategy can show that $P \& Q$ but P cannot explain Q is consistent with physicalism.

(iii) The knowledge gap: Q cannot be deduced from P

Before I prove that the phenomenal concept strategy can show how the knowledge gap is consistent with physicalism, I will have to argue that the knowledge gap is simply another version of the conceptual gap. The canonical argument for the knowledge gap, Jackson's Mary in the black-and-white room, has us conceive of the situation in which Mary is in her black-and-white room with knowledge of P . Then we're told to conceive of Mary leaving to room, and asked whether Mary would gain new knowledge of Q . Which is to say: is it conceivable that Mary gains new knowledge of Q ? The end result is that we're conceiving of Mary as having knowledge of P but not knowledge of Q , in a world different from ours where P and Q also happen to be true.

How should we represent the truth-involving knowledge gap in this case? (To save room, let's introduce a knowledge operator $K()$ such that " $K(P)$ " means "one has knowledge of P ") The truth-neutral version of the knowledge gap would be: $P \& Q \& K(P) \& \sim K(Q)$ is conceivable. Now, of course, to make this truth-neutral knowledge gap a truth-involving gap will require adding $P \& Q$ outside of the conceivable scenario. So the truth-involving knowledge gap should be represented as: $P \& Q$ and $K(P) \& P \& \sim K(Q) \& Q$ is conceivable.

Given P , Q , and C^* , the phenomenal concept strategy can show how the knowledge gap is consistent with physicalism. Because the entire thought experiment is conceptual, it must be the case that $K(P)$ and P are both physical facts conceived under physical concepts, while $\sim K(Q)$ and Q are both phenomenal facts conceived under phenomenal concepts. If C^* is true then there is a conceptual disconnect between $K(P)\&P$ and $\sim K(Q)\&Q$, and so $K(P)\&P\&\sim K(Q)\&Q$ is conceivable. So given P , Q , and C^* , the phenomenal concept strategy can show that $P\&Q$ but $K(P)\&P\&\sim K(Q)\&Q$ is conceivable is consistent with physicalism.

(3) C^* is physically explicable. On all the phenomenal concept strategy accounts reviewed in Section 4, phenomenal concepts are said to refer to phenomenal states in an intimate and essential manner without revealing anything about their referents' nature that lends itself to physical or functional analysis. As a result, the accounts are entirely compatible with physicalism. Moreover, unlike phenomenal experiences themselves, there is nothing about phenomenal concepts attributed to us by C^* that would make them incompatible with physicalism. Some argue that physical accounts can only explain two things: spatiotemporal structure and function in terms of causal roles played in the production of a system's behavior. While it seems very likely that explaining structure and function will not suffice to explain phenomenal experience, what makes the phenomenal concepts attributed to us by C^* so unique is that they can be characterized in terms of structural and functional features. Therefore, unlike phenomenal experience itself, there is good reason to think that C^* is physically explicable, or at least has the potential for further scientific exploration.

What I've just done is constructed a general representation of the phenomenal concept strategy as putting forward a thesis C^* , and shown that (1) C^* explains the conceptual disconnection between physical concepts and phenomenal concepts, (2) Given $C^* \& P \& Q$, and the CE Principle, the phenomenal concept strategy can show how the relevant truth-involving epistemic gaps are consistent with physicalism, and (3) C^* is physically explicable. Let's now run it through Chalmers' master argument to make sure it is immune.

9. Chalmers' master argument, reprise

The new phenomenal concept strategy with C^* should work for the type-B physicalist and avoid Chalmers' second-order explanatory gap problem. Chalmers claims that not only do conscious beings and their zombie-counterparts have corresponding sets of beliefs, they also have corresponding sets of concepts (although not necessarily with the same referents). Intuitively, it is plausible that a zombie's physical concepts and phenomenal concepts are no less conceptually connected than a conscious being's concepts.

But on top of this, we should recognize that C^* makes a claim about what scenarios are conceivable, and evaluate whether C^* will be true of zombies based on what is conceivable to a zombie.

First, because both conscious beings and their zombie-counterparts have corresponding beliefs and corresponding concepts, they should be able to correspondingly conceive of the same things we can. There's no reason to think that a zombie who can have the belief Q cannot also conceive of $P \& Q$ as well as $P \& \sim Q$.

Second, the notion of conceivability we've been using, which says that the conceivability of S requires that the truth of S cannot be ruled out a priori, gives us no reason to believe that consciousness would be necessary for conceivability. So long as zombies have corresponding conceptions of the same three scenarios we can conceive of $(P \& Q, P \& \sim Q, \sim P \& Q)$, I'm inclined to say that C^* is true of zombie-world as well.

The upshot is that the proponent of the phenomenal concept strategy respond to Chalmers' dilemma by saying that $P \& \sim C^*$ is inconceivable given the assumption that C^* can be physically explained. As a result C^* will be true of zombies. This says that zombies must be able to conceive of the same things as us, which they can.

I believe that I have successfully shown how my representation of the phenomenal concept strategy avoids Chalmers' master argument while showing how the relevant truth-involving epistemic gaps are consistent with physicalism. My conclusion is that the anti-physicalist inference to an ontological gap has been undercut in the case of the three relevant epistemic gaps.

References

- Aydede, M. & Güzeldere, G. (2005). "Cognitive architecture, concepts, and introspection: An information-theoretic solution to the problem of phenomenal consciousness", *Nous*, 39 (2): 197-255.
- Balog, K. (1999). "Conceivability, Possibility, and the Mind-Body Problem", *Philosophical Review* 108: 497-528.

- Balog, K. (2006). "Acquaintance and the Mind-body Problem", <http://pantheon.yale.edu/~kb237/Web%20publications/Acquaintance.pdf>, (unpublished), accessed 2009.
- Balog, K. (2009a). "Phenomenal Concepts", *The Oxford Handbook in the Philosophy of the Mind*, Oxford: Oxford University Press.
- Balog, K. (2009b). "In Defense of the Phenomenal Concept Strategy", *Philosophy and Phenomenological Research* (forthcoming).
- Block, N. (2006). "Max Black's Objection to Mind-Body Identity", *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, ed. T. Alter and S. Walter. Oxford: Oxford University Press.
- Carruthers, P. (2004). "Phenomenal Concepts and Higher-Order Experiences", *Philosophy and Phenomenological Research*, 68:316-36.
- Carruthers, P. & Veillet, B. (2007). "The Phenomenal Concept Strategy", *Journal of Consciousness Studies* 12.
- Chalmers, D. (1996). *The Conscious Mind*, New York: Oxford University Press.
- Chalmers, D. & Jackson, F. (2001). "Conceptual Analysis and Reductive Explanation", *Phil Review* 110:315-61.
- Chalmers, D. (2002a). "Consciousness and Its Place in Nature", *Philosophy of Mind: Classical and Contemporary Readings*, Oxford: Oxford University Press.
- Chalmers, D. (2002b). "Does Conceivability Entail Possibility?" *Conceivability and Possibility*, ed. T. Gendler & J. Hawthorne. Oxford: Oxford University Press, pp.145-200.
- Chalmers, D. (2003). "The Content and Epistemology of Phenomenal Belief", *Consciousness: New Philosophical Perspectives*, ed. Q. Smith and A. Jokic, Oxford: Oxford University Press, 220-72.
- Chalmers, D. (2007). "Phenomenal Concepts and the Explanatory Gap", *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, ed. T. Alter and S. Walter. Oxford: Oxford University Press.
- Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. Cambridge: MIT Press
- Fodor, J. (1990). *A Theory of Content and Other Essays*. Cambridge: MIT Press.
- Hill, C. (1997). "Imaginability, Conceivability, Possibility, and the Mind-Body Problem", *Philosophical Studies*, 87:61-85
- Hill, C. & McLaughlin, B. (1999). "There are Fewer Things in Reality than are Dreamt of in Chalmers's Philosophy", *Philosophy and Phenomenological Research*, 59:445-54.
- Ismael, J. (1999). "Science and the Phenomenal" *Philosophy of Science*, 66:351-69.
- Kripke, S. (1972). *Naming and Necessity*. Cambridge: Harvard University Press.

- Levin, J. (2007). "What is a Phenomenal Concept?", *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, ed. T. Alter and S. Walter. Oxford: Oxford University Press, pp.87-110.
- Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.
- Levine, J. (2007). "Phenomenal concepts and the materialist constraint", *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, ed. T. Alter and S. Walter. Oxford: Oxford University Press.
- Loar, B. (1990/97). "Phenomenal states", *Philosophical Perspectives*, 4:81-108. Revised edition in (ed. N Block, O. Flanagan and G. Güzeldere.) *The Nature of Consciousness*. Cambridge, Mass: MIT Press.
- Millikan, R. (1989). "Biosemantics", *Journal of Philosophy*, 86:281-97.
- Nagel, T. (1974). "What Is It Like to Be a Bat?" *Philosophical Review*, 83:435-50.
- O'Dea, J. (2002). "The Indexical Nature of Sensory Concepts", *Philosophical Papers*, 31:169-81
- Papineau, D. (1993). *Philosophical Naturalism*. Oxford: Blackwell.
- Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.
- Papineau, D. (2007). "Phenomenal and Perceptual Concepts" *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, ed. T. Alter and S. Walter. Oxford: Oxford University Press pp.111-44.
- Perry, J. (2001). *Knowledge, Possibility and Consciousness*, Cambridge, Mass: MIT Press
- Russell, B. (1910). "Knowledge by acquaintance and knowledge by description", *Proceedings of the Aristotelian Society*, 11:108-128, Reprinted in: *Russell, Mysticism and Logic*. London: Allen and Unwin, 1963, 152-167
- Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge: MIT Press.
- Stoljar, D. (2005). "Physicalism and phenomenal concepts", *Mind and Language*.
- Sturgeon, S. (1994). "The Epistemic Basis of Subjectivity", *Journal of Philosophy*, 91: 221-35.
- Tye, M. (2003). "A Theory of Phenomenal Concepts", *Philosophy*, 53:91-2016
- Van Inwagen, P. (1997). "Modal Epistemology", *Philosophical Studies*, Netherlands: Kluwer Academic Publishers.