



Conceptual Analysis and Reductive Explanation

Author(s): David J. Chalmers and Frank Jackson

Source: *The Philosophical Review*, Jul., 2001, Vol. 110, No. 3 (Jul., 2001), pp. 315-360

Published by: Duke University Press on behalf of Philosophical Review

Stable URL: <https://www.jstor.org/stable/2693648>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



Duke University Press and are collaborating with JSTOR to digitize, preserve and extend access to *The Philosophical Review*

JSTOR

Conceptual Analysis and Reductive Explanation

David J. Chalmers and Frank Jackson

1. Introduction

Is conceptual analysis required for reductive explanation? If there is no a priori entailment from microphysical truths to phenomenal truths, does reductive explanation of the phenomenal fail? We say yes (Chalmers 1996; Jackson 1994, 1998). Ned Block and Robert Stalnaker say no (Block and Stalnaker 1999).

A number of issues can be distinguished:

- (1) Is there an a priori entailment from microphysical truths to ordinary macroscopic truths?
- (2) If there is no a priori entailment from microphysical truths to phenomenal truths, does reductive explanation of the phenomenal fail?
- (3) If there is no a priori entailment from microphysical truths to phenomenal truths, is physicalism about the phenomenal false?
- (4) Is there an a priori entailment from microphysical truths to phenomenal truths?

We hold that the first three questions should be answered positively (with some qualifications to be outlined). Block and Stalnaker hold that the first three questions should be answered negatively. Their central strategy is to argue for a negative answer to the first question, and to use this conclusion to argue for a negative answer to the second and third questions. They argue that truths about water and about life, for example, are not entailed a priori by microphysical truths, but that this is no bar to the reductive explanation and physical constitution of water and of life.

In this paper, we will address Block and Stalnaker's arguments for a negative answer to the first three questions, while remaining neutral on the fourth. We will proceed by first giving an independent defense of a positive answer to the first question. This makes the ensuing reply to Block and Stalnaker more straightforward, and also makes the discussion accessible to those unfamiliar with the literature.

Thanks to Ned Block and to two *Philosophical Review* referees for comments.

2. Clarifying the Thesis

The initial thesis at issue is whether there is an a priori entailment from microphysical truths to ordinary macrophysical truths.¹ Before we proceed, we need to clarify and qualify this thesis to meet some obvious objections.

On our usage, P entails Q when the material conditional $P \supset Q$ is true: that is, when it is not the case that P is true and Q is false. An a priori entailment is just an a priori material conditional. For ease of usage, we will speak of a priori entailment as *implication*. On this usage, P implies Q when the material conditional $P \supset Q$ is a priori; that is, when it is possible to know that P entails Q with justification independent of experience. On this usage, entailment is a nonmodal notion, while implication involves an epistemic modality. We will assume that there are some a priori truths; general skepticism about the a priori is beyond the scope of this paper.

Let P be the conjunction of microphysical truths about the world. Microphysical truths are truths about the fundamental entities and properties of physics, in the language of a completed physics. We can also stipulate that P includes the conjunction of the fundamental laws of physics. We will not engage the issue of what counts as “physics,” but will stipulate that if there are any fundamental mental entities or properties, they are not part of physics. (This begs no important questions: if the mental is fundamental, it cannot be reductively explained.) We also will not engage issues arising specifically from the foundations of quantum mechanics, and will take it for granted that microphysical truths include or imply truths about the spatiotemporal position, velocity, and mass of microscopic entities.² So P will likely include or imply truths about the distribution of fundamental entities (perhaps particles, waves, and/or fields) in space-time, truths about their fundamental properties, and truths about the fundamental laws that govern them.

Let M be a typical macroscopic truth concerning natural

¹Horgan (1984) puts forward a version of this thesis under the name “cosmic hermeneutics,” although he qualifies the thesis to allow a role for a posteriori identities involving rigid designators in inferring macroscopic truths from microscopic truths. Chalmers (1996), Jackson (1994, 1998), and Lewis (1994) argue for versions of the thesis.

²To simplify, we can assume something like a Bohmian interpretation of quantum mechanics. The central claim can also be defended under other interpretations, but further complexities are involved.

phenomena, such as water or life, outside the psychological, social, and evaluative domains. Some such truths include: 'Water boils at 100 degrees Centigrade', 'water is H_2O ', 'water covers much of this planet', 'life propagates through the replication of DNA', 'there are many living beings', and so on. The initial thesis at issue is that for this sort of M , P implies M . This thesis needs to be immediately qualified to handle some obvious loopholes.

(1) *Negative truths*. As described, P is compatible with the claim that there are further nonphysical entities or properties in the world: angels, perhaps, or epiphenomenal ectoplasm. So P does not imply such negative truths (if they are truths) as the truth that there are no angels, or that there is no epiphenomenal ectoplasm. For similar reasons, P does not imply such universally quantified truths as the truth (if it is a truth) that all living beings contain DNA molecules.

This loophole can be closed by conjoining to P a "that's-all" statement T , asserting that our world is a *minimal* world satisfying P . Intuitively, this statement says that our world contains what is implied by P , and *only* what is implied by P . More formally, we can say that world $W1$ *outstrips* world $W2$ if $W1$ contains a qualitative duplicate of $W2$ as a proper part and the reverse is not the case. Then a minimal P -world is a P -world that outstrips no other P -world.³ It is plausible that no world containing angels is a minimal P -world: for any P -world containing angels, there is an angel-free P -world that it outstrips. So $P \wedge T$ implies that there are no angels. For similar reasons, $P \wedge T$ implies

³It is well known that (just as in the case of defining intrinsic properties) some notion in this circle has to be taken as primitive: perhaps the notion of minimality, or of outstripping, or of containing a qualitative duplicate. One also might invoke the notion of intrinsic property itself, for example, to say that world $W1$ outstrips if all the intrinsic properties and relations in $W2$ are correspondingly instantiated in $W1$, and if the reverse is not the case.

A complication arises in characterizing the that's-all claim in terms of minimality where a priori entailment, rather than necessitation, is concerned. The relevant claim must hold that the actual world is minimal among the class of *epistemic* possibilities satisfying P , where an epistemic possibility corresponds intuitively to a maximally specific hypothesis that is not ruled out a priori. On some philosophical views, this might come apart from a definition in terms of minimality among the class of metaphysical possibilities. For example, on a view on which it is necessary but not a priori that God exists, 'God exists' might be entailed by P conjoined with an assertion of metaphysical minimality, but not conjoined with an assertion of epistemic minimality. This can be handled straightforwardly by formalizing the notion of epistemic possibility (e.g., Chalmers forthcoming b). Alternatively, where this sort of case is concerned, one can retain the intuitive characterization of the that's-all statement.

the other negative and universally quantified statements mentioned above. So the thesis at issue should be that for the relevant macroscopic truths M , $P \wedge T$ implies M . (From now on we will abbreviate ' $P \wedge T$ ' as ' PT ', and similarly for other conjunctions among the four statements discussed in this section.)

(2) *Indexical truths*. As described, the information contained in PT specifies a world objectively. For this reason, it does not imply any indexical truths: truths such as 'I am Australian', or 'life evolved in the past', or 'there is water on this planet'. It also may not imply such truths as 'water is made of H_2O '. If the universe contains an inhabited planet where there is a superficially identical liquid made of XYZ , then the information in PT alone will not allow us to decide whether we live on the H_2O planet or on the XYZ planet, so PT is at least epistemically compatible with the claim that water is made of XYZ .

This loophole can be closed by conjoining to PT some *locating* information (or *indexical* information) I . I can be thought of as a "you are here" marker added to the objective map given by PT . I can consist of the conjunction of any two truths 'I am A ' and 'now is B ', where A is an identifying description of myself (or the subject in question) and B is an identifying description of the current time. (An identifying description is a description such that PT implies that there is a unique individual or time satisfying the description.) As long as PT implies all relevant objective information, then I will enable the subject to "locate" himself or herself, and PTI will imply all indexical truths.⁴ So the thesis at issue should be that for the relevant macroscopic truths M , PTI implies M .

(3) *Phenomenal truths*. One of the questions at issue is whether phenomenal truths—truths concerning states of phenomenal consciousness—are implied by microphysical truths. Many hold that they are not, and hold that adding indexical and that's-all information does not close the gap. We will not adjudicate that issue here. But we note that if phenomenal truths are not implied by PTI , then it is likely that many other macroscopic truths are not so implied either. For example, knowing whether an object is red arguably requires knowing whether it is the sort of object that causes a certain sort of color

⁴Strictly speaking, to handle some cases (for example, the "Two Tubes" case discussed in Austin 1990), I needs to contain further indexical information, such as information about the referent of certain special demonstratives (see Chalmers forthcoming 2002a). This subtlety can be passed over for present purposes.

experience, and knowing whether an object is hot arguably requires knowing whether it is the sort of object that causes experiences of heat. If so, then if truths about color experience and heat experience are not implied by *PTI*, truths about color and heat are not implied either.

This loophole can be closed while remaining neutral on the important issues, by conjoining *Q*, the conjunction of all phenomenal truths, to *PTI*. *Q* will specify the phenomenal states and properties instantiated by every subject bearing such states and properties, at every time.⁵ That is, for every subject who is phenomenally conscious at a given time, *Q* will specify precisely what it is like to be that subject at that time. We can also stipulate that *Q* includes any fundamental principles that lawfully govern phenomenal states or that connect phenomenal states with physical states. The thesis at issue is then whether *PQTI* implies the relevant macroscopic truths *M*, which we can now understand to include truths about color and heat, as well as phenomenal truths.

To see that this is a neutral way of posing the issue, note that if *PQTI* implies all such *M* (as we shall argue), then either (i) *PTI* alone implies all such *M*, or (ii) *PQTI* implies all such *M*, but *PTI* does not. In the first case, the original thesis involving *PTI* will be upheld. (In this case, phenomenal truths will themselves be implied by *PTI*.) In the second case, all failures of *PTI* to imply a relevant *M* will be associated with the failure of *PTI* to imply *Q* in the sense that adding *Q* will close any epistemic gaps. Putting these cases together, the thesis concerning *PQTI* entails the crucial claim that *if* phenomenal truths are not implied by *PTI*, then there is a special epistemic gap in the phenomenal case.

(If we combine these alternatives with the thesis that reductive explanation goes along with a priori entailment, then the first alternative above leads to the view that all the relevant macroscopic truths, including phenomenal truths, are reductively explainable. The second alternative leads to the view that phenomenal truths are not reductively explainable, and that other macroscopic truths are

⁵The specification of phenomenal states in *Q* should use terms that express what Chalmers (forthcoming 2002a) calls “pure phenomenal concepts,” picking out phenomenal states directly in terms of phenomenal character, as opposed to relational phenomenal concepts, which pick out phenomenal states in terms of their typical external causes or effects. We will be neutral here about just what is involved in phenomenal character.

reductively explainable “modulo phenomenology”—that is, that we can reductively explain those aspects of them for which the phenomenal plays no conceptually constitutive role. This would fit nicely with the view (articulated by Nagel 1974 and Searle 1991, among others) that the actual reductive explanation of phenomena such as color and heat has proceeded by explaining their objective aspects while leaving their subjective aspects untouched.)

Note that the addition of Q to the conjunction changes the way that we define T and I . T will now hold that the actual world is a minimal world satisfying PQ , and I will involve identifying descriptions of a subject and a time relative to the information in PQ . (In fact, if P does not imply Q and Q is true, then the original version of PTI is false.) So the crucial thesis is that $PQTI$, understood as the conjunction of P and Q with T and I so understood, implies M , for the relevant M . Note also that the information in I depends on the specification of a subject and a time, so that strictly speaking, $PQTI$ will vary between subjects and times, and the thesis will be that $PQTI$ implies M for any given subject and time in our community. In what follows, we will assume that an arbitrary subject and time have been selected.

3. A Priori Entailment and Conceptual Analysis

It is sometimes claimed that for $A \supset B$ to be a priori, the terms in B must be definable using the terms of A . On this view, a priori entailment requires definitions, or explicit conceptual analyses: that is, finite expressions in the relevant language that are a priori equivalent to the original terms, yielding counterexample-free analyses of those terms. This is not our view. The falsity of the claim can be seen from the following.

Let G be the conjunction of the statements in the following passage: ‘Smith believes with justification that Jones owns a Ford. Smith initially has no beliefs about Brown’s whereabouts. Smith forms a belief that Jones owns a Ford or Brown is in Barcelona, based solely on a valid inference from his belief that Jones owns a Ford. Jones does not own a Ford, but as it happens, Brown is in Barcelona’. Let K be the statement ‘John does not know that Jones owns a Ford or Brown is in Barcelona’.

It is plausible that $G \supset K$ is a priori. But it is also plausible that there is no explicit analysis of the concept of knowledge using the terms involved in G . If so, a priori entailment does not require explicit

analyses of the terms in the consequent using the terms in the antecedent. It is also somewhat plausible that there is no explicit analysis of the concept of knowledge at all. If so, a priori entailment does not require explicit analyses of the terms in the consequent.

The two important features of the case are the apriority of the conditional $G \supset K$ and the absence of explicit analyses of the concept of knowledge. On the first point: once general skepticism about the a priori is set aside, this conditional seems to be a typical example of an a priori truth. Someone who knows that G is true and who possesses the concepts involved in K (in particular the concept of knowledge) is thereby in a position to know that K is true, even if they lack any further relevant empirical information. That is, a grasp of the concept of knowledge (along with a grasp of the other concepts involved) and rational reflection suffice to eliminate the possibility that G is true and K is false.

This conditional plays an essential role in Gettier's argument for the conclusion that knowledge is not justified true belief. Gettier's argument was an a priori argument, in which empirical information played no essential role, and its conclusion is a paradigmatic example of a non-obvious a priori truth. The argument proceeds by presenting the possibility that G holds, and appealing to the reader's concept of knowledge to make the case that if G holds, K holds (and J holds, where J is a corresponding positive claim about John's justified true belief). Empirical information plays no essential role in justifying belief in this conditional, so the conditional is a priori. The a priori conditional itself plays an essential role in deriving the a priori conclusion.

On the absence of explicit analyses of knowledge: We take it that this is a reasonable conclusion to draw from four decades of failed attempts to produce explicit analyses. Certainly no explicit analysis has met with widespread approval, and proposed analyses are always confronted quickly by plausible counterexamples. Of course it could be that there is a correct explicit analysis that has not yet been produced, or that has been produced but overlooked. But even if so, it seems clear that the a priori entailment from G to K is not *hostage* to an explicit analysis of knowledge that would support the entailment. Whether or not there is such an analysis, the entailment is a priori all the same.

So a priori entailment does not require explicit analysis. If anything, the moral of the Gettier discussion is that the reverse is

often true: explicit analyses are themselves dependent on a priori intuitions concerning specific cases, or equivalently, on a priori intuitions about certain conditionals. The Gettier literature shows repeatedly that explicit analyses are hostage to specific counterexamples, where these counterexamples involve a priori intuitions about hypothetical cases. So a priori conditionals seem to be prior to explicit analyses at least in matters of explicit justification; and in general there is no reason to hold that a priori conditionals need explicit analyses to underwrite them.

It could be argued that while these a priori entailments are not underwritten by explicit analyses, they are underwritten by explicit *sufficient* conditions for knowledge or its absence—for example, the condition that a belief solely based on inference from a false belief is not knowledge. Of course it is trivial that there is a sufficient condition in the vicinity of such an entailment (the antecedent provides one such), so the claim will be interesting only if the complete set of sufficient conditions for knowledge is not indefinite and open-ended. But the Gettier literature suggests precisely that the set of sufficient conditions for knowledge is open-ended in this way; if it were not, we would have a satisfactory explicit analysis. And as before, the a priori entailments are not hostage to the proposed sufficient conditions; if anything, proposed sufficient conditions are hostage to a priori intuitions about specific cases.

Once an essential role for explicit definitions is eschewed, the model of conceptual analysis that emerges is something like the following. When given sufficient information about a hypothetical scenario, subjects are frequently in a position to identify the extension of a given concept, on reflection, under the hypothesis that the scenario in question obtains. Analysis of a concept proceeds at least in part through consideration of a concept's extension within hypothetical scenarios, and noting regularities that emerge. This sort of analysis can reveal that certain features of the world are highly relevant to determining the extension of a concept, and that other features are irrelevant.

What emerges as a result of this process may or may not be an explicit definition, but it will at least give useful information about the features in virtue of which a concept applies to the world. It will usually be the case that one can find complex expressions whose conditions of application approximate those of the original concept to some degree, where one finds increasingly good approximations through increasingly complex expressions. In this way we can

elucidate at least many important aspects of how a concept's extension depends on the world. But in general, there is no reason to suppose that a finite expression yielding a counterexample-free analysis of a concept must result from this process.⁶ This pattern, whereby a conditional ability to evaluate a concept's extension yields elucidation of a concept without a finite counterexample-free analysis, is illustrated very clearly in the case of 'knowledge'.

The possibility of this sort of analysis is grounded in the following general feature of our concepts. If a subject possesses a concept and has unimpaired rational processes, then sufficient empirical information about the actual world puts a subject in a position to identify the concept's extension. For example, if a subject possesses the concept 'water',⁷ then sufficient information about the distribution, behavior, and appearance of clusters of H₂O molecules enables the subject to know that water is H₂O, to know where water is and is not, and so on. This conditional knowledge requires only possession of the concept and rational reflection, and so requires no further a posteriori knowledge.

Of course this claim is trivial if the empirical information in question is allowed to include information directly involving the concept at issue (for example, 'water'). But as this case and the case of knowledge suggests, the claim is often true even when that sort of information is excluded: a 'water'-free description of the world can enable one to identify the referent of 'water', and a 'knowledge'-free description of the world can enable one to decide whether a given belief is an instance of knowledge. In these cases, we can say that *nontrivially* sufficient information enables identification of a concept's extension. We will not try to give a precise account of what counts as nontrivially sufficient information (presumably one should also exclude near-synonymous expressions, such as 'aqueous' or 'epistemic'), but will leave the notion intuitive. Note that there is no requirement at this point that the information in question be microphysical, or that it be about perceptual evidence, or that it be

⁶FJ is somewhat more optimistic than DJC about the possibility of satisfactory finite analyses, especially if one recognizes that conceptual analysis can accommodate an element of conceptual revision to clear up confusions in a folk concept. See Jackson, Pettit, and Smith 2000 for some reasons for optimism.

⁷We will use quoted expressions for both concepts and terms in this article, and will slide between the levels of thought and language for ease of discussion; nothing turns on this.

the information in *PQTI*.

This ability to identify a concept's extension is not restricted to true empirical information about the actual world. If the world had turned out differently, we could still have identified the concept's extension. Correspondingly, we can evaluate the concept's extension given *hypothetical* information about ways the actual world might be. Let us say that a hypothesis is epistemically possible (in the broad sense) when it is not ruled out a priori. Let us say that an *epistemically possible* hypothesis characterizing the total state of the world corresponds to an *epistemic possibility*: intuitively, a specific way the actual world might turn out to be, for all one can know a priori.⁸ Then sufficient information about an epistemic possibility enables a subject to know what a concept's extension will be, under the hypothesis that the epistemic possibility in question is actual. For example, in the Gettier case, it is irrelevant whether John's case is actual: a subject can know that *if* John's case as described is actual, then John does not know that someone owns a Ford. Or in the case of water, given appropriate information about the distribution, behavior, and appearance of clusters of XYZ molecules (information analogous to the information we have about H₂O in the actual world), a subject is in a position to conclude that *if* the information is correct, then water is XYZ.

If something like this is right, then possession of a concept such as 'knowledge' or 'water' bestows a *conditional ability* to identify the concept's extension under a hypothetical epistemic possibility, given sufficient information about that epistemic possibility and sufficient reasoning. That is, possession of these concepts in a sufficiently rational subject bestows an ability to evaluate certain conditionals of the form $E \rightarrow C$, where E contains sufficient information about an epistemic possibility and where C is a statement using the concept and characterizing its extension, for arbitrary epistemic possibilities.⁹ And conceptual analysis often proceeds precisely by evaluating conditionals like these.

In the most important cases, these conditionals will be a priori. Certainly there will be related cases in which $E \rightarrow C$ is a posteriori: for

⁸See Chalmers (forthcoming b) for a formalization of this notion. The informal characterization suffices for present purposes.

⁹Determinate application of these conditionals may be restricted to epistemic possibilities that are not too far from home. When epistemic possibilities deviate greatly from our presuppositions about the actual world, some of our concepts will lose determinate application.

example, it is a posteriori that if a glass contains H_2O , it contains water. But these will be cases in which the antecedent E does not contain sufficient empirical information to identify the concept's extension given possession of the concept alone. The a posteriority of these conditionals reflects the fact that further empirical information is required for their justification. But then all we need to do is conjoin E with the relevant further empirical information F , and we will obtain a conditional $E' \supset C$ that is knowable a priori, where E' is the conjunction of E and F . For example, in the case of 'water', identification of the concept's extension requires a great deal of further information about the distribution, behavior, and appearance of clusters of H_2O molecules in the world. But once this information F is conjoined with the original information E , we obtain a more complex conditional of the form $E' \supset C$ that is plausibly a priori.

Given sufficient empirical information in the antecedent, there is good reason to believe that the resulting conditionals will be a priori. These a priori conditionals will reflect the way in which we can identify a concept's referent. If we possess a concept, then sufficient empirical information E enables us to conclusively identify the concept's extension and to know associated truths C , and we are in a position to do this whether the information in E is actual or hypothetical. Because all the relevant empirical information is present in the antecedent of the conditional, empirical information plays no essential role in justifying belief in the conditional. So $E \supset C$ is a priori. We can call this sort of conditional an *application conditional*.

We do not claim that application conditionals with nontrivially sufficient information in the antecedent exist for all concepts. For example, it may be that there is no way to know truths about time without having empirical information that more or less explicitly invokes the concept of time. If so, there will be no antecedents with nontrivially sufficient information to imply consequents about time. Some other primitive concepts (space? cause? consciousness?) may also be like this. But for many or most concepts, there will exist application conditionals (corresponding to arbitrary epistemic possibilities) whose antecedents contain nontrivially sufficient information.

It should be noted that nothing here conflicts with the conclusions of Kripke (1972) about names and natural kind terms. On Kripke's picture, it is an a posteriori necessity that water is H_2O . But this is entirely compatible with there being an a priori conditional from

certain (false) statements about the distribution, behavior, and role of XYZ to 'water is XYZ'. Kripke allows that a conditional can be a priori even when it is not necessary. For example, his view allows that claims such as 'heat (if it exists) is the dominant cause of heat sensations' may be a priori, although not necessary. The same goes for many claims with similar form, such as claims about the length of the meter stick in Paris. If so, then many conditionals of the same form as 'if *X* is such-and-such, then *X* is heat' may be a priori, even where it is an a posteriori necessity that heat is not *X*.¹⁰

One can put this by making the familiar observation that even if it is not metaphysically possible that water is XYZ, it is epistemically possible that water is XYZ, in the broad sense that it is not ruled out a priori that water is XYZ. It is also epistemically possible that XYZ has a certain specific distribution, behavior, and appearance (of a sort that is characteristic of water). A subject possessing the concept of 'water' can reason straightforwardly that *if* the second hypothesis obtains (in the actual world), then the first hypothesis obtains. This sort of a priori relationship among epistemic possibilities is entirely compatible with different necessary relationships among metaphysical possibilities.

There is also nothing here that contravenes Kripke's epistemological arguments against certain descriptive views of reference-fixing. Indeed, these arguments can be seen as invoking just this sort of reasoning about epistemic possibilities. In considering whether the term 'Gödel' is a priori equivalent to 'the prover of the incompleteness of arithmetic', Kripke considers a certain epistemically possible hypothesis on which the incompleteness of arithmetic was proved by a man named 'Schmidt' and was then stolen and

¹⁰Some contemporary philosophers (e.g., Salmon (1986)) go beyond Kripke's view, holding that statements such as 'heat (if it exists) is the dominant cause of heat sensations' (and the like) are not a priori, and that many apparently empirical identities of the form '*X* is *Y*' are a priori. This counterintuitive view is usually held for theory-dependent reasons. For example, if a philosophical theory holds that statements like these express singular propositions (so that '*X* is *Y*' expresses the same singular proposition as '*Y* is *Y*'), and that the apriority of a statement is a function of the proposition it expresses, then given certain assumptions about the apriority of singular propositions the theory will be committed to these counterintuitive claims. We think that this is itself a good reason to reject the theories in question, and that there are also strong independent reasons to reject such theories, but we cannot go into that matter here. For present purposes, we will simply assume that the intuitive claims about apriority are correct.

promulgated by a friend named 'Gödel'. Kripke suggests that if this hypothesis is actual, then our term 'Gödel' refers not to Schmidt, who proved incompleteness, but to his friend. If so, it is not a priori that Gödel (if he existed) proved the incompleteness of arithmetic. This is a straightforward example of armchair reasoning about how the extension of a concept depends on how the actual world turns out.

One observation must be made in the case of names. It is possible that two people who use a given name might use it with *different* a priori application conditionals. For example, if Leverrier uses 'Neptune' as a name for whatever planet perturbs the orbit of Uranus, then conditionals of the form 'If X is the planet that perturbs the orbit of Uranus, X is Neptune' will be knowable a priori by him. But if Leverrier's wife uses 'Neptune' knowing only that it is an astronomical object for which her husband is searching, then the conditional just mentioned will not be knowable a priori by her. She will still have an ability to know the term's extension given sufficient information about the actual world, but for her, the antecedent information will crucially require information about her husband's intentions. So possession of 'Neptune' gives the wife and the husband conditional abilities to apply the term that are somewhat different from each other: they yield different application conditionals, and they yield a different pattern of application across epistemic possibilities.

This sort of epistemic variability suggests that at least in the case of names, we should take the apriority of a sentence to be subject-relative. For example, we can say that a sentence S is a priori for a speaker when the belief that S would express for the speaker can be justified independently of experience, yielding a priori knowledge.¹¹ So above, 'Neptune (if it exists) perturbs the orbit of Uranus' is a priori for Leverrier but not for his wife. The relevant class of a priori conditionals will then be subject-relative in a similar way. But it remains the case that when a subject possesses a name, the subject will have a conditional ability to identify its extension given sufficient empirical information about the actual world, and the relevant

¹¹Or we might say: a sentence S is a priori for a subject when an assertion of S by that subject would express an a priori justifiable thought, where thoughts are the sort of propositional attitude tokens apt to be directly expressed by an assertion, including beliefs and belief-like attitudes such as entertainings. The move from beliefs to thoughts accommodates the possibility that the subject does not believe what he or she asserts. An a priori justifiable thought is a thought that can be justified independently of experience, on idealized rational reflection, yielding a priori knowledge.

conditionals will be a priori for the subject.

The case of names is incidental to our central purposes, but it is arguable that something similar can occur for a natural kind term such as ‘water’. The most obvious cases of this will occur when a subject uses the term with deference to others in a linguistic community. Such cases will be not unlike the case of Leverrier’s wife above. For present purposes, it is probably best to take it that we are stipulating that the terms in question are used nondeferentially. Deference raises issues that are orthogonal to the central issues here about reductive explanation. Nothing important to those issues would change if we were dealing with a community of one speaker, or with only “expert” subjects.

It may be that a natural-kind term can be epistemically variable for reasons independent of deference. For example, perhaps a city-dweller might use ‘water’ nondeferentially for the liquid that comes out of faucets (knowing nothing of oceans), and a beach-dweller might use ‘water’ nondeferentially for the liquid in the oceans (knowing nothing of faucets). This sort of case can be treated as we suggested that cases involving names be treated: the subjects have different conditional abilities, and different associated conditionals will be knowable a priori for each of them. But as before, this epistemic variability gives no reasons to deny that the a priori conditionals exist. In what follows, we will usually abstract away from this sort of cross-subject variability, as the central issues about reductive explanation are largely orthogonal to these questions about variable use within a community. But when necessary, all of the claims about apriority in the following can be put in subject-relative terms.

4. On the Entailment of Macroscopic Truths

We can now address the crucial question: For the relevant macroscopic truths M , does $PQTI$ imply M ?

We have already made the case that for many such truths M , there is *some* nontrivially sufficient empirical information E such that E implies M . The question now is whether the specific empirical truth $PQTI$ is such an E —that is, whether $PQTI$ contains sufficient information to imply M .

A conclusive argument for this thesis would require very detailed discussion, but here we will present some reasons for finding the thesis plausible. The basic argument has a two-step structure. First,

PQTI implies complete information (in the language of physics) about the structure, dynamics, composition, and distribution of macroscopic systems, as well as information about the actual and potential perceptual appearances that they present. Second, this information about macroscopic structure, dynamics, composition, distribution, and appearance (along with residual information from *PQTI*) implies ordinary macroscopic truths such as *M*. But we can take things a little more slowly. To address the question, it is useful to start by imagining that one starts with *only* the empirical information specified by *PQTI*, and by asking, Could one thereby come to know the truth of *M*?

To start with, one can get some distance using *Q* alone. On a phenomenalist view, *Q* alone (construed as the complete truth about actual and counterfactual experiences) implies all truths. Even if phenomenalism is rejected, *Q* still gives a significant epistemic foothold on the world. Combining the complete phenomenal information in *Q* with the indexical information in *I* puts me in a position to determine the phenomenal character of my current experiences, and of my experiences throughout my lifetime. This will include in particular the phenomenal character of a lifetime of *perceptual* experiences. This information serves as at least an epistemic guide to many macroscopic truths, just as it does in ordinary life. If *V* is a specific phenomenal character of a visual experience as of a large object in front of me, then if I know that I am now having an experience with phenomenal character *V*, then I might reasonably infer that there is a large object in front of me.¹² The same goes for many other perceptual experiences.

Of course, this information does not remotely suffice to imply all the relevant macroscopic truths *M*. First, there is a vast class of truths about which my perceptual experience gives no guidance: truths concerning unperceived objects, for example. Second, for those truths *M* about which my perceptual experience gives guidance, it

¹²Note that as before, *Q* will characterize the experience according to its phenomenal character alone, leaving causal relationships to the environment unspecified. So *Q* will not characterize *V* as the phenomenal character of visual experiences caused by large objects, and knowledge of *Q* will not build in this sort of causal knowledge. Rather, knowledge of *Q* will build in knowledge of phenomenal character—that is, knowledge of what it is like to have the relevant experiences. The relevant point here is that knowledge of phenomenal character alone plausibly gives some epistemic guidance about the nature of the environment, even if it leaves open many skeptical hypotheses.

remains epistemically possible (in the broad sense) that I have these perceptual states but that *M* is false. Such epistemic possibilities range from traditional skeptical scenarios concerning the nonexistence of the external world to a wide range of scenarios involving perceptual illusions, false testimony, false abductive or inductive inferences, and so on.

Nevertheless, truths about which my perceptual experience gives no guidance can often be settled by further information; and skeptical scenarios can often be ruled out by further information. Starting with the further information in *Q*, information about others' experiences and about counterfactual experiences will give guidance about many unperceived objects. Of course numerous skeptical scenarios will still be left open, at least if phenomenalism and related views are rejected. But once the information in *P* (and *T* and *I*) is admitted, many such scenarios are immediately ruled out: those in which the physical world does not exist, for example. The residual question is whether the information in *PQTI* suffices to derive knowledge of the unknown truths, and to rule out all the skeptical scenarios.

The information in *P* will play a crucial role. This includes complete information about the structure and dynamics of the world at the microphysical level: in particular, it includes or implies the complete truth about the spatiotemporal position, velocity, and mass of microphysical entities. This information suffices in turn to imply information about the structure and dynamics of the world at the macroscopic level, at least insofar as this structure and dynamics can be captured in terms of spatiotemporal structure (position, velocity, shape, etc.) and mass distribution. For example, for any given region of space at a time, the information in *P* implies information about the mass density in the region, the mass density in various subregions, the causal connections among various complex configurations of matter in the region, and the extent to which the matter in the region behaves or is disposed to behave as a coherent system. This information suffices to determine which regions are occupied wholly by causally integrated systems that are disposed to behave coherently. So the information plausibly suffices for at least a geometric characterization—in terms of shape, position, mass, composition, and dynamics—of systems in the macroscopic world.

The central point here is that a macroscopic description of the world in the language of physics is implied by a microscopic

description of the world in the language of physics. Such a thesis is extremely plausible: it is not subject to any worries about translation between vocabularies, and involves only a change of scale. The only worry might concern the status of bridging principles within physical vocabulary: for example, is it *a priori* that the mass of a complex system is the sum of the masses of its parts? If there are any concerns here, however, they can be bypassed by stipulating that the relevant physical principles are built into *P*.¹³ *P* also implies information about systems' microstructural composition, and about their distribution of systems across space and time, including the relations between systems (characterized in macrophysical terms) and about any given system's history (characterized in macrophysical terms).

Further, the information in *P* and *Q* together will imply truths about regularities connecting the physical and phenomenal domains. *PQTI* will include or imply all truths about lawful covariation between the world's physical and phenomenal states. If my own phenomenal states depend directly on associated physical states, then *P* and *Q* in conjunction with *I* will imply information about the dependence. If there are certain regularities by which other physical systems in the world indirectly affect my perceptual phenomenal states, then *P*, *Q*, and *I* will imply information about those regularities. So although information about the external causes of perceptual phenomenal states is not built in to *Q*, *PQTI* will imply information about these causes. It will also imply information about the perceptual phenomenal states that various external systems are disposed to cause when appropriately situated: that is, about the perceptual appearance of these systems.

Overall, *PQTI* implies complete information about the (geometrically characterized) structure and dynamics of macroscopic systems and objects in the world, their spatiotemporal distribution and microstructural composition, and their actual and potential perceptual appearances. This information puts a subject in a position to conclusively know (on rational reflection) the truth or otherwise of any ordinary macroscopic claim *M*. Complete knowledge of

¹³In fact, for present purposes one could even stipulate that *P* contains a complete description of the world, both microscopic and macroscopic, in the language of physics. The central issues about consciousness, and the central issues that divide us from Block and Stalnaker, arise equally whether one starts with microphysics or macrophysics, and so are unaffected by such a stipulation. Still, on our view such a stipulation is unnecessary.

perceptual appearances yields the information that members of our community rely on in coming to know macroscopic truths; and complete structural, dynamic, distributional, and compositional information contains all the information that we need to settle the truth of claims that perceptual information does not settle.

For example, knowledge of the appearance, behavior, and composition of a certain body of matter in one's environment, along with complete knowledge of the appearance, behavior, and composition of other bodies of matter in the environment, and knowledge of their relationships to oneself, puts one in a position to know (on rational reflection) whether the original system is a body of water. The same goes for knowledge of whether the system is gold, whether it is alive, whether it boils at a certain temperature, or whether it is found in the oceans. And the same applies to ordinary macroscopic truths *M* in general: complete knowledge of structure, dynamics, composition, distribution, and appearance puts one in a position to know whether *M* is true.

Further, the information in *PQTI* serves to conclusively eliminate arbitrary skeptical hypotheses under which *M* might be false. Hypotheses involving perceptual illusions or hallucinations are eliminated by full structural and dynamical information. Hypotheses concerning differences in the past and the future are eliminated by full distributional information. Hypotheses concerning differences in underlying causal or compositional structure are eliminated by full compositional information. Even skeptical hypotheses concerning differences in others' minds are plausibly eliminated by full phenomenal information. Further skeptical hypotheses about the subject's own relation to these systems, or about their exhaustiveness, are removed by the indexical and that's-all information in *PQTI*. A relevant skeptical hypothesis would have to be one in which the structure, dynamics, distribution, composition, and appearance of objects and systems across space and time is preserved (along with indexical and that's-all information), but on which *M* is false. There do not seem to be any such: the knowledge outlined above suffices for conclusive knowledge of *M*.

Importantly, no other empirical information is required to justify the inference from *PQTI* to *M*: *PQTI* contains all the information that is needed to know *M*. We can imagine a subject engaged in a Cartesian suspension of all empirical belief, who is then given the information that *PQTI*. Given this information alone, the subject

would be in a position to reconstruct knowledge of the structure, dynamics, distribution, composition, and appearance of external systems by the reasoning above, and from there to ascertain the truth of macroscopic claims such as M and to eliminate any relevant skeptical hypotheses. All that is required here is possession of the concepts in M , the information in $PQTI$, and rational reflection.

So knowledge of $PQTI$ suffices in principle for conclusive knowledge of M , with no other empirical information required. The same reasoning applies hypothetically: so even without knowing $PQTI$, a subject is in a position to know that if $PQTI$ is true, then M is true, and to rule out any hypothesis on which $PQTI$ is true but M is not. So the subject is in a position to know the truth of the material conditional $PQTI \supset M$.¹⁴ As before, empirical information plays no essential role in justifying knowledge of this conditional (all the information required is present in the antecedent), so the subject is in a position to know the conditional a priori.¹⁵ It follows that $PQTI$ implies M .

As before, a priori knowledge of $PQTI \supset M$ does not rely on any explicit analysis of the concepts involved in M , or on any explicit bridging principles connecting the vocabulary of $PQTI$ with the vocabulary of M . Just as a 'knowledge'-free description of a Gettier situation implies relevant claims about knowledge without requiring an explicit bridge between the vocabularies, $PQTI$ implies the truth of M without requiring an explicit bridge between the vocabularies. Rather, $PQTI$ serves as sufficient information for determining the truth of M , in the sense described earlier. In effect, this breaks down into two stages: $PQTI$ serves as sufficient information for determining complete information about structure, dynamics, distribution, composition, and appearance, and this information serves in turn as sufficient information for determining the truth of M . So $PQTI \supset M$ is an a priori application conditional.

¹⁴One might worry: if M involves a natural kind term, might belief in truths involving the term (even conditional truths) require acquaintance with the relevant kind? In response: even if acquaintance is required for possession of the concept and so for the relevant belief, it does not enter essentially into the belief's justification, for the reasons above. Whatever one says about acquaintance, it is plausible that a subject competent with the terms can in principle use ' $PQTI \supset M$ ' to express a belief, such that the belief is true, justified, and constitutes knowledge, and such that empirical information does not enter essentially into the justification of this knowledge. So $PQTI \supset M$ will be a priori for the subject.

¹⁵A more detailed argument against empirical justification is given in section 5, part (6).

It might be objected that we do not yet have a completed physics, so that we do not yet know what *P* says, so we cannot know whether *PQTI* implies *M*.¹⁶ But for present purposes, we do not need to know what *P* says. All we have needed here is that *P* implies truths about the structure and dynamics of macroscopic objects in spatiotemporal terms, along with truths about mass and about microphysical composition. And that implication is likely to be robust over physical theories. (If a physical theory does not yield this implication, we are unlikely to count it as a complete physical theory.) In any case, we stipulated earlier that we are assuming a physical theory that implies truths about the position, velocity, and mass of all microscopic entities. Given this assumption, the step to an analogous characterization of macroscopic systems proceeds as outlined above.

It might also be objected that no human could grasp all the information in *PQTI*, so that no human could grasp the truth of the relevant conditional. This is surely true, but it is no bar to the apriority of the conditional. Apriority concerns what is knowable in principle, not in practice, and in assessing apriority, we idealize away from contingent cognitive limitations concerning memory, attention, reasoning, and the like. Once we idealize away from human memory and processing limitations, the problem here is removed.

We can summarize the argument in a way that makes clear the role of the idealization, by appeal to a fanciful thought experiment. Imagine a human augmented by a “virtual world” machine. This is a machine containing (i) a supercomputer to store the physical information in *P* and to make the a priori calculations required to move from microscopic structure to macroscopic structure, (ii) a virtual reality device to produce direct knowledge of the phenomenal states described in *Q*, and (iii) tools that use these devices to focus on arbitrary regions of the world, and to deliver information about the macroscopic structure, dynamics, composition, and perceptual appearance of systems in those regions. Using such a machine, a

¹⁶Byrne (1998) makes this sort of objection against an argument for a priori entailment in Chalmers 1996. Byrne’s objections in this paper pass over the two-step character of the entailment, from microphysical structure and dynamics to macrophysical structure and dynamics, and from there (in combination with perceptual appearances) to the relevant macroscopic truths. (This two-step structure is present in Chalmers 1996, albeit briefly.) His discussion also passes over the role of epistemological considerations regarding the elimination of skeptical scenarios.

human with no other empirical information could straightforwardly ascertain the truth of the relevant claims M . The virtual world does no more than give access to the information contained in $PQTI$ and process this information on a priori grounds. So if a human using such a device can ascertain the truth of M , it is plausible that $PQTI \supset M$ is a priori.

One might ask: given that $PQTI$ implies ordinary macroscopic truths, does similar reasoning suggest that PTI alone implies many such truths? Such reasoning may work for truths concerning the objective spatiotemporal structure of macroscopic systems. But for many or most macroscopic concepts (most obviously concepts such as 'green' and 'hot', but plausibly also concepts such as 'water' and 'tiger'), our application of the concepts relies essentially on associated perceptual appearances. For a typical truth M involving such a concept, PTI will imply M only if PTI implies truths about the associated perceptual appearances. And this will plausibly require that PTI implies the relevant phenomenal truths in Q .

One can then ask: could reasoning such as the above establish that PTI implies the truths in Q ? It seems not. There are familiar reasons (e.g., Jackson 1982) to hold that information about the structure, dynamics, and composition of physical systems does not suffice for information about the character of conscious experiences. And there are familiar reasons (e.g., Chalmers 1996) to hold that there are conceptually coherent alternative possibilities in which things are physically just as in PTI , but in which different sorts of experiences are present, or in which there are no experiences at all.¹⁷ Whether or not these reasons are sound, their existence makes it clear that there is no direct argument that PTI implies Q analogous to the argument concerning $PQTI$ above. Any argument that PTI implies Q must rest on quite different considerations.¹⁸

Note that we have not argued here that $PQTI$ implies every truth in every domain (although we are inclined to accept this claim). Given what we have said here, it could be that certain truths in special domains—perhaps concerning mathematics, metaphysics, morality,

¹⁷A reviewer asks whether this can be squared with knowledge of other minds. Yes: just as the information in Q can be an epistemic guide to truths about the external world without implying those truths, the information in P (and information about one's own mind) might be an epistemic guide to truths about other minds without implying those truths.

¹⁸FJ articulates considerations of this sort in forthcoming work.

or mentality?—are not implied by *PQTI*.¹⁹ Presumably this will be because these truths are not conclusively knowable even given full information about structure, dynamics, appearance, distribution, composition, and so on. Perhaps these truths are knowable only through different means, or perhaps they are not conclusively knowable at all. But we hope we have said enough to make it plausible that ordinary macroscopic truths concerning everyday macroscopic natural phenomena are implied by *PQTI*.

5. Block and Stalnaker on A Priori Entailment

Block and Stalnaker give a number of arguments against a priori entailments from microphysical truths to macroscopic truths. They do not always cast things in these terms: their discussion often suggests that the issue is whether there is an analysis of macroscopic concepts “in microphysical terms,” and whether principles that cross the bridge from microphysical to macroscopic are themselves “microphysical claims.” This way of putting things is unfortunate, as on a natural reading these theses are implausible and much stronger than what is needed. But it is clear from the context of Block and Stalnaker’s discussion that a priori entailment is the real issue, and many of their arguments are naturally seen as arguments against a priori entailments.

(1) *Explicit analyses*. One aspect of the global structure of Block and Stalnaker’s article can be seen as follows. In the first part of their article (especially sections 4–9), they argue there are no explicit analyses of typical macroscopic concepts (such as ‘water’ and ‘life’) of a sort that could support an a priori entailment from microphysical to macroscopic truths. In the second part of their article (the long section 10), they discuss the use of the two-dimensional semantic framework as an alternative to explicit analyses, and argue that the existence of the framework does not support the a priori entailments

¹⁹One might worry in particular about mental truths concerning subjects’ propositional attitudes. If these truths are not implied by *PQTI*, then it is likely that many other truths will not be: for example, truths involving social concepts (for example, ‘money’) or involving terms used deferentially (for example, ‘arthritis’, used by a non-expert). For a more general thesis that applies to such truths, one would need to argue that *PQTI* implies these mental truths (as we hold), or to expand *PQTI* to include mental truths of a more general sort. This question does not affect the present discussion, as truths involving ordinary macroscopic concepts of the natural world (such as ‘water’ or ‘gold’, used nondeferentially) do not seem to be tied to propositional attitudes in this way.

in question. The intended upshot is that the existence of these a priori entailments is doubtful, and that reductive explanation does not require a priori entailment.

An argument with this structure cannot successfully make a case against a priori entailments. To see this, it suffices to note that if this sort of argument succeeds, it succeeds equally in making a case against the a priori Gettier entailments discussed earlier. It is at least as plausible for 'knowledge' as for 'water' and for 'life' that there is no explicit analysis to support the entailments. And the general criticisms of this use of the two-dimensional framework presumably apply equally to its use in the case of 'knowledge'. But nevertheless, the a priori Gettier entailments discussed earlier exist; or at least, it is clear that this sort of argument does little to make a case against them. So by parity, this sort of argument does little to make a case against the a priori entailments we are concerned with.

We can agree with Block and Stalnaker that there are plausibly no precise explicit conceptual analyses of concepts such as 'water' and 'life' (or at least, no precise analyses of a manageable length). But as we made clear earlier (and as is also made clear in Chalmers 1996 and Jackson 1998), such explicit analyses are not required for a priori entailment.²⁰

We can also agree with Block and Stalnaker that the mere existence of the two-dimensional semantic framework does not imply that the a priori entailments in question exist. As Block and Stalnaker say, the two-dimensional framework provides a good way to capture a priori connections when they exist, but the framework alone does not demonstrate specific a priori connections. But the two-dimensional framework plays no essential role in our arguments for a priori entailments. It played no role in the arguments given here. In the

²⁰As part of their argument, Block and Stalnaker suggest that the most plausible conceptual analyses are restricted to connections within the same "family" of terms (as with the analysis of 'bachelor' as 'unmarried man'), so that analyses of macroscopic concepts in microphysical terms are implausible. Ned Block has suggested in correspondence that this point extends to a priori entailments, so that the only plausible a priori entailments involve connections within a family. The reasons for accepting this thesis are unclear, however. The main evidence for it seems to lie in the fact that the most familiar and uncontroversial a priori entailments involve intrafamily connections. On the view we are advocating, however, this evidence is easily explained by the fact that there are few *short* cross-family a priori entailments, since cross-family entailments will usually have detailed and unwieldy antecedents.

arguments given in Chalmers (1996) and Jackson (1998), it plays only a clarifying role, in removing certain confusions that may arise from the presence of a posteriori necessary connections, and in providing a convenient shorthand for discussing the patterns by which a concept applies to the world. The core of the argument for a priori entailments proceeds independently of these semantic details, just as it does in the Gettier case.²¹

Insofar as it is captured by this global structure, then, Block and Stalnaker's discussion does not engage the first-order issue of whether the a priori entailments in question exist. But Block and Stalnaker's article also contains numerous arguments that go beyond what is suggested by this structure. We will discuss these in what follows.

(2) *Conceivability*. In a discussion about conceivability, Block and Stalnaker discuss the issue (raised by Levine 1993) of whether it is conceivable that P holds without W holding, where P is the complete microphysical truth and W is 'water is boiling'. They discuss two senses of conceivability. In the first sense, conceivability is tied closely to a posteriori possibility and necessity, so that $P \wedge \sim W$ is inconceivable; but this sense of conceivability is clearly irrelevant to the issue of a priori entailment. In the second sense, P without Q is conceivable if Q cannot be deduced from P using logic and conceptual truths. This sense is more clearly relevant to the issue; we would prefer that the notion be cast in the more general terms of a priori reasoning rather than explicit deduction, but this does not matter to Block and Stalnaker's discussion.²² They say:

Let C be a complete description, in microphysical terms, of a situation in which water (H_2O) is boiling, and let T be a complete theory of physics.

²¹For this reason, we will not address Block and Stalnaker's discussion of the two-dimensional framework in this paper. On our use of the two-dimensional framework, the primary intension (or A-intension) of a concept reflects the a priori conditionals that capture the way a concept applies within various epistemic possibilities, in the manner suggested in section 3 of this paper. This understanding of the framework in epistemic terms differs subtly from Block and Stalnaker's understanding in terms of context-dependence; for discussion, see Chalmers (forthcoming d).

²²Block and Stalnaker's first sense of conceivability is a version of what Chalmers (forthcoming 2002b) calls "secondary conceivability." Their second sense is a version of "primary conceivability," and in particular is close to Chalmers' "ideal primary negative conceivability." Block and Stalnaker note that this notion is a purely negative notion, so that the name "conceivability" is misleading. It is worth noting that there is also a closely related positive notion in the vicinity, which Chalmers calls "ideal primary positive conceivability."

Can one deduce from *T*, supplemented with analytic definitions, that H_2O would boil in circumstances *C*? To see that one cannot, suppose that the deduction is taking place on Twin Earth. The stuff they call 'water' is XYZ, and the process they call 'boiling' is a process that superficially resembles boiling, but that involves a different physical process. Just as they would say (truly), "Water is XYZ, and not H_2O (and if there were H_2O , it wouldn't be water)," so they would say (truly), "If there were H_2O , and it were behaving like that, it wouldn't be boiling." They could hardly deduce ' H_2O would boil in circumstances *C*' if on their meaning of 'boil', H_2O can't boil at all. (We assume that boiling is a natural kind concept. If you don't agree, substitute some other process term that does express a natural kind concept.) (8)

We can agree with Block and Stalnaker that if there are residents of Twin Earth, they might truly say "If there were H_2O , and it were behaving like that, it wouldn't be boiling." But this is irrelevant to the issue of whether there is an a priori entailment from ' H_2O is behaving like that' to ' H_2O is boiling'. The first is a subjunctive conditional about an explicitly counterfactual scenario, and as such is relevant to issues concerning necessity, not apriority.

Compare the following: There is plausibly an a priori entailment from something like 'molecular motion is the dominant cause of our heat sensations' (*M1*) to 'molecular motion is heat' (*M2*). But on a Twin Earth, where heat sensations are caused by something very different, residents might truly say, "If molecular motion were the dominant cause of our heat sensations, it wouldn't be heat." The latter is a subjunctive claim about an explicitly counterfactual scenario, so at best its truth reflects negatively on the necessity of an entailment from *M1* to *M2*. Its truth clearly has no bearing on the apriority of the entailment (which remains plausible for everything said here). The same goes for microphysics and the boiling of water.

(In fact, the thought experiment does not even bear directly on the necessity of the entailment, since presumably the necessity of an utterance in English does not entail the necessity of the homophonic utterance in Twin English (witness 'water is H_2O '). If anything, the thought experiment bears on whether it is a priori that the entailment is necessary.)

Block and Stalnaker also say:

We don't really need a Twin Earth story to make our point. Consider a person on actual Earth, who does not know the story about how water boils—perhaps she doesn't even know that water is made up of molecules. One presents her with the theory *T*, and a description (in

microphysical terms) of a water boiling situation. Can she then deduce that if T is true and a situation met conditions C , then the H_2O would be boiling? No, since for all she knows the actual situation is like the one on Twin Earth. Perhaps, if she were told, or could figure out, that the theory was actually true of the relevant stuff in her environment, she could then conclude (using her knowledge of the observable behavior of the things in her environment) that H_2O is water, and that the relevant microphysical description is a description of boiling, but the additional information is of course not a priori, and the inference from her experience would be inductive. (8)

This example is simply an illustration of the general phenomenon whereby physical information must be supplemented by locating information (“you are here”) to imply relevant conclusions. The subject needs locating information to know that the theory describes her own environment and not a very different environment. So while this example may show that there is no implication from P (the complete physical truth) to W , it does nothing to show that there is no implication from PI (physical truth supplemented by locating information) to W . Of course (as is made clear in Chalmers 1996 and Jackson 1998) the latter is the more relevant claim.

(3) *Uniqueness*. In discussing the possibility of conceptual analyses of terms like ‘water’, Block and Stalnaker spend a considerable amount of time discussing the possibility of analyzing this term as something like ‘the waterish stuff around here’, where ‘waterish’ abbreviates a cluster of descriptions of familiar macroscopic properties of water (such as clarity, liquidity, drinkability, and so on). They argue that no such analysis can succeed. The argument proceeds by considering various possible cases where there is more than one sort of stuff in the environment that satisfies the relevant descriptions. Given that we hold that explicit analysis is not required for a priori entailment, we could simply set this discussion aside as irrelevant, but there are some points that are worth discussing.

Block and Stalnaker suggest at one point that for an a priori entailment to go through, water must be analyzed as something like ‘the *unique* waterish stuff in our environment’. This seems clearly false. First, it seems obviously epistemically possible (in the broad sense) that there could turn out to be two sorts of stuff with the relevant properties filling the oceans and lakes in our environment, in roughly equal distributions. In such a case, it seems clear that we would say that there are two sorts of water (that is, this epistemically

possible hypothesis is an instance of the epistemically possible hypothesis that there are two sorts of water). So any a priori analysis requiring uniqueness would clearly be incorrect.

Second, the falsity of an analysis involving uniqueness does nothing to rule out a priori entailment. The thought experiment above itself involves largely a priori consideration of an epistemically possible hypothesis. A priori consideration of that hypothesis suggested that it counts as an instance of the epistemically possible hypothesis that there are two sorts of water. So the thought experiment is entirely compatible with an a priori entailment from specific details concerning the hypothesis (in 'water'-free terms) to a conclusion involving water, and indeed may even positively suggest the existence of such an entailment.

If one wanted to come up with an explicit analysis that handles these cases directly, one might start simply with 'the watery stuff in our environment', which handles them well (where the phrase is naturally interpreted to cover all instances of watery stuff in our environment). There will be cases where this analysis does not work: perhaps if there is a single dominant watery stuff and small pockets of different stuff, we might count only the dominant stuff as 'water'. We could then move to something like 'the dominant watery stuff in our environment, if there is one; or any watery stuff in our environment, if there is not'. But no doubt this analysis would itself need to be refined to handle more complex cases, just as in the case of knowledge. It is probably easier, then, to give up the aim of producing a perfect explicit analysis, and to content ourselves with the observation that we have an a priori grasp of how our concepts apply to specific epistemic possibilities, when these are described in sufficient detail. Nothing about these cases contravenes this observation.

(4) *Ghost heat*. In the course of their discussion of uniqueness, Block and Stalnaker discuss a hypothetical epistemic possibility in which it turns out that there is "ghost heat." In this epistemic possibility, things are microphysically just as normal (P is true), so the causal role associated with heat is filled by molecular motion. But the causal role is also filled by a nonphysical substance, which we can call 'ghost heat', in a case of causal overdetermination. Block and Stalnaker say that under this epistemic possibility, it turns out that there are two sorts of heat (molecular motion and ghost heat), so that 'heat = molecular motion' turns out to be false. So P does not imply that heat is molecular motion.

We are not certain that this sort of causal overdetermination is coherent, and we are not certain that if it is coherent, then the epistemic possibility in question should be interpreted as one in which there are two kinds of heat. But even if so, the conclusion is simply an instance of the failure of the microphysical facts P to rule out epistemic possibilities in which there are further nonphysical facts. As before, ruling out these epistemic possibilities requires conjoining a “that’s-all” statement T , holding that the actual world is a minimal P -world to P . Once T is conjoined to P (yielding PT), the hypothesis involving ghost heat is ruled out, as a world with ghost heat is clearly not a minimal P -world. So the example shows at best that P does not imply H (where H is ‘heat is molecular notion’); it does not show that PT does not imply H . And (as is clear in Chalmers 1996 and Jackson 1998), the latter is the more relevant claim.

Block and Stalnaker anticipate this sort of response, and reply:

Recall that Jackson’s definition of physicalism recognized the possibility of a world that is a microphysical duplicate of the actual world, while also containing some additional nonphysical substances and properties. In response, a “nothing but” condition is built into the definition of physicalism by requiring for the truth of physicalism only that *minimal* physical duplicates of the actual world be duplicates simpliciter. But the need for this qualification presupposes that the “nothing but” condition is not something that is entailed by microphysics itself; that is, *it is not a claim of microphysics that our world is a minimal physical duplicate of itself*. To take the “nothing but” condition to be an implicit claim of microphysics would be to *build the thesis of physicalism into microphysics*, which philosophers such as Jackson and Chalmers, who reject physicalism, should be reluctant to do. They reject physicalism, not microphysics. The truth of what physicists write in textbooks does not depend on the mind-body problem. Even if it is a microphysical fact that H_2O is *a* waterish stuff around here, it is not a microphysical fact that it is *the* waterish stuff around here. (19)

The objection here seems to be that the “that’s-all” statement T is not itself implied by P . This is correct but irrelevant. The thesis is not that T is part of microphysics. The theses are rather that (i) PT (perhaps in conjunction with QI) implies the relevant macroscopic truths; that (ii) this sort of implication is required for reductive explanation; and that (iii) an analogous claim (that PT necessitates all truths) is required for physicalism. Nothing that Block and Stalnaker say here gives any reason to reject these claims.

(5) *Methodology and simplicity*. In the background of Block and

Stalnaker's specific arguments about uniqueness, there lie some more general worries about the role of a posteriori methodological considerations, such as simplicity, in determining the extension of our terms. These worries are not elaborated at length in Block and Stalnaker's discussion, but they come up in a number of places and in a number of different ways.

For example, in the discussion of ghost heat, Block and Stalnaker suggest that it is by using considerations of simplicity that we rule out the possibility of ghost heat. That seems roughly correct. Our observations are compatible with two hypotheses: $P \wedge T$ and $P \wedge G$, where P contains the relevant physical truths, T is a "that's-all" statement, and G is a claim about ghost heat. We do not decide between these hypotheses purely through observation, or through deduction, but through the use of simplicity considerations such as Ockham's razor. Something similar goes on in inductive inference, and in abductive inference, and so on. These considerations play a common role in moving from partial information about the world to more complete information about the world.

All this is quite compatible with our position. It is no part of our position that a posteriori methods are never used in determining our concepts' extensions. They are used all the time: in particular, it is by using a posteriori methods that we form empirical hypotheses about how the world is. The information in $PQTI$ itself is knowable only a posteriori; and simplicity considerations play a central role in our coming to know aspects of this information. But nothing here bears on the claim that *given* the information about $PQTI$ (which rules out the $P \wedge G$ hypothesis), the application of our concepts is determined a priori.

There is one passage that may bear directly on the question above. After considering our intuitions about how the term 'water' might apply to various epistemic possibilities, including possibilities in which there are multiple substances that play the relevant role, Block and Stalnaker write:

It is a part of the semantics of natural kind terms that they are natural kinds, but it may also be part of the semantics of these terms that this is a defeasible condition. What is not plausibly part of the semantics, something we all know in virtue of knowing our language alone, is what to say in all the myriad cases in which the defeasible condition is defeated. In these cases, what we should say will no doubt be dictated by principles of "simplicity," conservativeness, etc.

Block and Stalnaker do not say anything to back up these claims

directly, or to argue against an opponent who holds that these conditions of application to various epistemic possibilities are as much a (tacit) part of the concept of 'water' as the conditions of application brought out in the Gettier literature are a (tacit) part of the concept of 'knowledge'. It should also be noted that even if these conditions of application are not part of the semantics of 'water' in English, this does not entail that a subject's application of the term to epistemic possibilities is not justified *a priori*. As we saw above in the case of 'Neptune', it may be the case that the relevant conditionals involving a term may vary between users of a term (so that the corresponding conditions of application are not built in to the term's semantics in English), but that each user's knowledge of the conditionals is justified *a priori* all the same. So Block and Stalnaker need to give substantive arguments against the *a priori* knowability of the conditionals in question.

Block and Stalnaker go on to discuss the case of 'rheumatism', in which as scientists acquired more empirical information about various diseases underlying a central syndrome, the term began to be used for the syndrome rather than the diseases. And they discuss the case of 'jade', in which we discovered two underlying substances and decided that there are two sorts of jade. Nothing here contravenes the claim that these decisions manifested a tacit grasp of *a priori* conditions of application of the terms in question, and that the relevant conditionals (from a full specification of the empirical details to conclusions about jade and rheumatism) are *a priori*.

Of course, the claim should not be made too strong. There is almost certainly a high degree of indeterminacy in our concepts, in their application both to the actual world and to hypothetical epistemic possibilities. It can sometimes happen that when an epistemic possibility is found to be actual, no clear decision about the concept's application is dictated. In such a case there often ensues (explicit or implicit) terminological stipulation or refinement, influenced by various sociological and pragmatic factors. It may be that in the 'rheumatism' case, for example, the initial extension of the concept was indeterminate between applying to the syndrome in general, and applying to the most common underlying disease. If so, there is an equally rational possible history in which the term comes to be applied to the disease; but in our world, it came to be applied to the syndrome.

Such cases can naturally be seen as cases of mild conceptual change, or of conceptual refinement. In these cases, a term that formerly expressed one concept comes to express a slightly different concept, so that (i) the term's extension becomes more determinate, (ii) certain sentences whose truth values were previously indeterminate come to express truths, and (iii) certain conditionals that were not *a priori* come to be *a priori*. But these changes happen in parallel, and there is no time at which a concept's extension is not determined *a priori* by truths about the underlying state of the world.

Conceptual change is the exception rather than the rule. Most of the time, decisions concerning a concept's extension in response to empirical developments are dictated rationally rather than arbitrarily. Cases of this sort pose no threat to the *a priori* entailment thesis. To see this, note that in these cases, the relevant empirical developments could be presented in advance as an epistemically possible hypothesis, and decisions concerning the concept's extension under that hypothesis will themselves be rationally accessible in principle. So the empirical knowledge in question plays no essential role in justifying the conditional from the empirical developments to a conclusion about the concept's extension. So in these cases we have no argument against *a priori* entailment.

(6) *A priori knowledge versus armchair knowledge.* A more general worry may be lurking in the background. It may be that Block and Stalnaker are prepared to concede that in many cases, a conditional from a sufficiently complete description of an epistemic possibility to a conclusion about a concept's extension is knowable "from the armchair," but they may deny that it is *a priori*. This view is suggested by the following passage:

This seems to be armchair reasoning, reflection that does not include any obvious reference to real experiments, so it is tempting to conclude that this reflection just unfolds our concepts in a totally *a priori* way. But what this conclusion misses is that our reasoning about the proper epistemic response in various counterfactual situations is informed not only by our concepts, but by implicit and explicit theories and general methodological principles that we have absorbed through our scientific culture—by everything that the "we" who are performing these thought experiments believe. What people should rationally say in response to various hypothesized discoveries will vary depending on their experience, commitments and epistemic priorities. (43)

On this view, armchair knowledge of application conditionals depends in part on general empirical background knowledge. For

example, the empirical background knowledge that the world is simple may play a role in our requiring that the extension of a concept be as simple as possible. Or perhaps the knowledge that the world contains numerous natural kinds may make us more inclined to apply terms in the manner typical of natural kinds. The view can be put as follows. For relevant truths M : (i) a subject can know an application conditional $PQTI \supset M$ from the armchair; (ii) there is some general empirical information E , acquired at some time in the past, that plays a role in justifying the subject's knowledge of the conditional; and (iii) empirical knowledge such as E plays an *essential* role in justifying knowledge of the conditional. If (i) and (ii) are correct, the conditional is not known a priori; if (iii) is also correct, the conditional is not knowable a priori.

We argued in section 4 that empirical knowledge E does not play this sort of justificatory role. But it is useful to examine further the roles that empirical information might play in knowledge of application conditionals. There is no question that empirical information can play a *causal* role in acquiring this knowledge. Empirical knowledge often plays a causal role in the acquisition of concepts with certain a priori connections, and it sometimes plays a role in triggering changes in the a priori connections associated with a term, as in the cases above. There is also no question that E could play a *mediating* role in our knowledge of an application conditional. That is, E might itself be implied by $PQTI$, and one could then straightforwardly use E in combination with $PQTI$ to deduce M . But neither of these possibilities entails that E plays an essential role in *justifying* knowledge of the relevant conditionals. We suspect that the cases in which Block and Stalnaker think that empirical information has a justificatory role are in fact cases where the information has a causal or mediating role.

To examine this matter in the context of the current objection, we will take it for granted that considerations earlier in the paper establish that for empirical truths M , the conditional $PQTI \supset M$ is at least knowable from the armchair. The same considerations suggest that the truth or falsity of conditionals of the form $X \supset M$ will be knowable from the armchair, for (false) antecedents X representing specific epistemic possibilities that are analogous to but distinct from $PQTI$. (Such possibilities might include different configurations of physical and phenomenal properties, and possibilities involving different laws of

nature.) Further, the earlier considerations suggest that the relevant armchair knowledge is sufficient to justify hypothetical reasoning from the hypothetical acceptance of X to the conclusion that M . (Note that this entails that our knowledge of these conditionals is not essentially grounded, uninterestingly, in armchair knowledge that X is false.) We can call the relevant conditionals *armchair conditionals*. Now the question is: Insofar as an empirical claim E plays a role in knowledge of armchair conditionals such as $PQTI \supset M$, does it play a justificatory role or merely a causal or mediating role?

To adjudicate this issue, we can focus on the status of E itself. By the usual reasoning, knowledge of $PQTI$ will enable a subject to know from the armchair whether E is true. For example, the information in $PQTI$ puts us in a position to know that the world is reasonably simple.²³ So there will be corresponding armchair conditionals of the form $PQTI \supset E$. Further, since E is a posteriori, there will also be epistemic possibilities (in the broad sense) under which E is *false*: for example, epistemic possibilities under which the world is not simple. So there will be antecedents X (for example, describing a non-simple world) such that $X \supset \sim E$ is an armchair conditional. Although the epistemic possibility specified by X falsifies E , we can nevertheless evaluate it using various ordinary concepts, yielding corresponding armchair conditionals $X \supset C$.

We can then ask: Does the role played by E in knowledge of armchair conditionals $PQTI \supset M$ extend to a role in knowledge of armchair conditionals $X \supset C$ where X falsifies E ? (For example, does our knowledge that the world is simple play a role in our application of concepts to a non-simple epistemic possibility?) If E plays a causal role in knowledge of application conditionals, we would expect the answer to be yes. If E plays a mediating role, we would expect the answer to be no. If E plays a justifying role, it is not clear what we would expect the answer to be. In fact, one can argue that whether the answer is yes or no, E plays no essential role in justifying the application conditionals.

If the answer is yes: then for such an X and C , $X \supset C$ and $X \supset \sim E$

²³ An opponent might try to avoid this line of argument by holding that the truth or falsity of E cannot be settled even by armchair reasoning from antecedents such as $PQTI$ and X . This position would require a very different sort of argument, however, arguing directly against the considerations in favor of armchair knowledge in section 4. The position is also implausible in Block and Stalnaker's favored case of simplicity.

are armchair conditionals, so the subject can reason hypothetically from the information that X to the conclusion that C and $\sim E$. So if the subject were to discover that X were actual (assuming rationality), the subject would come to know that C and $\sim E$. E would then play no essential role in justifying the subject's knowledge that C , as the justification for items of knowledge cannot be essentially grounded in a falsehood. A subject who rationally evaluates X as a hypothetical can come to conclude C (under that hypothesis) by the same rational process; so E plays no essential role in justifying acceptance of the conditional from X to C . So in this application of the concepts involved in C , E plays no essential justificatory role. At most, E plays a causal role in the subject's acquiring a concept with this pattern of application.

If the answer is no: then the justificatory role of E extends at most to armchair conditionals $Y \supset C$ such that there is armchair justification for $Y \supset E$. Further: if there is armchair justification for $Y \supset E$, then if E justifies $Y \supset C$, $Y \supset E$ will also justify $Y \supset C$ without antecedently assuming E (by hypothetical reasoning from Y to E to C). So justification of $Y \supset E$ is required for E -involving knowledge of $Y \supset C$, and given this justification, E plays no further essential role in justifying $Y \supset C$. In particular, justification of $PQTI \supset E$ is required for E -involving knowledge of $PQTI \supset C$, and given this justification, E plays no further essential role in justifying $PQTI \supset C$. This suggests strongly that in these cases, E plays only a mediating role in knowledge of $PQTI \supset C$.

The only residual question about this second class of cases is what justifies the armchair conditional $PQTI \supset E$. Knowledge of this conditional is not essentially justified by E itself. To see this, note that a subject who is given the information in $PQTI$ and who antecedently suspends judgment about E will nevertheless be in a position to conclude that E holds. Further, the same sorts of factors are plausibly relevant in justifying $PQTI \supset E$ and in justifying $X \supset \sim E$ for those X that falsify E ; but E cannot justify the latter, for the reasons given above. Alternatively, $PQTI \supset E$ might be justified by some other empirical factor F ; but then knowledge of E does not play an essential role in justifying $PQTI \supset C$, contrary to the original hypothesis. Knowledge that F is all that is required; and we can run the same argument for F . The same goes for the disjunction of E and F , and indeed for any empirical claim. So no empirical claim plays an

essential role in justifying knowledge of armchair conditionals of the form $PQTI \supset C$. At most, such claims play a mediating role.

To illustrate the situation, we can take the case of simplicity, and the example of $PQTI \supset M$. On both alternatives described here, a posteriori simplicity considerations might play some role in adjudicating the status of M . On the first alternative, this is because these considerations have played a causal role in the subject's possessing a concept whose a priori conditions of application across all epistemic possibilities involve simplicity. On the second alternative, it is because the information that the world is simple is itself derivable from $PQTI$, and thereby plays a role in the a priori process of reasoning from antecedent to consequent. Either way, the a posteriori considerations play no role in justifying the conditional itself, and are therefore no bar to its apriority.

It might be objected that if empirical considerations can play a role in affecting what is a priori involving a concept (as in the first alternative above regarding simplicity, and also on the version of the 'rheumatism' case involving terminological stipulation), then the notion of apriority is being watered down. But this seems wrong: apriority is a matter of non-empirical *justification*. Concept acquisition is usually empirically driven, and conceptual drift can occur in response to empirical factors, but neither of these is any bar to the apriority of resulting claims involving the concepts.

A related objection notes that if the apriority of various sentences can be affected by empirical developments, then there is no guarantee that claims we now regard as a priori will turn out to be a priori; as a result, any metaphysical and explanatory conclusions that rest on claims about apriority are suspect. But this is wrong. Even where the sort of conceptual drift described here occurs, the relevant claims at the first stage are still a priori; it is just that later, the same sentence is used to express something different that is not a priori. The change here is merely terminological. Whether people use 'rheumatism' for the syndrome or the disease, the substantive issues concerning the phenomenon itself are the same; and whether people choose to use terms in a way that requires simplicity or not, the substantive issues concerning the phenomena are the same. The change in language simply means that we express things differently.

It may even be that in the future, people may come to use a term such as 'consciousness' or 'life' with a priori application conditions

that differ from ours, due to sociological or pragmatic factors, or terminological stipulation, or terminological drift. But this sort of future terminological change has no bearing on the truth or the apriority of claims that we currently make using the term, and has no bearing on any metaphysical or explanatory conclusions that might follow.²⁴ For example, it could turn out that due to this sort of drift, what someone later calls ‘consciousness’ can be reductively explained; but that does not imply that consciousness can be reductively explained.

We conclude that none of Block and Stalnaker’s arguments against *a priori* entailment are compelling.²⁵

6. Does Reductive Explanation Require *A Priori* Entailment?

Block and Stalnaker hold that reductive explanation does not require *a priori* entailment.²⁶ In arguing for this thesis, their main strategy involves arguing that in paradigm cases of reductive explanation, no *a priori* entailment exists. We have already rebutted these arguments. But they also offer a positive model of reductive explanation with some arguments in support. So it is worth saying something here about what reductive explanation involves. We start by examining a clear way in which reductive explanation can succeed and a clear way in which it can fail.

When a concept of some natural phenomenon supports *a priori* entailments from the microphysical, there is a clear sense in which the

²⁴This objection reflects a point that occasionally arises in discussion of the two-dimensional framework. Cannot the primary intension (or *A*-intension) of a term change, and if so, does this not call any conclusions drawn from *a priori* reasoning into question? In response: yes, the primary intension of a term can change (though this sort of change is not especially common and is usually minor); and no, this terminological change has no bearing on conclusions drawn from *a priori* reasoning with our current concept. A related but better objection holds that we can be wrong about the primary intension of a concept because we have not reasoned sufficiently deeply about what to say about a given epistemic possibility. This can certainly happen, but it does not cast doubt on the *a priori* method. Rather, it suggests that the *a priori* method has to be practiced well.

²⁵Yablo (2002) gives some further arguments against the case for *a priori* entailment, based partly on the role of “sensibility” in deriving macroscopic knowledge from microscopic knowledge. Chalmers (2002b) replies to these arguments.

²⁶A close tie between reductive explanation and *a priori* entailment is suggested by Chalmers 1996, Levine 1993, and Loar 1997.

phenomenon can be reductively explained. These a priori entailments might not support a *reduction* of the phenomenon in question to a microphysical phenomenon (at least in some senses of this term), perhaps because such entailments are compatible with multiple realizability. But nevertheless, in showing how any instance of the phenomenon is itself implied by microphysical phenomena, we show that there is a sort of transparent epistemic connection between the microphysical and macroscopic phenomena. Both the microphysical and the macroscopic phenomena are epistemically contingent, in that they involve the actualization of just one of a host of coherent epistemic possibilities. But where this sort of transparent entailment is present, the epistemic contingency in the macroscopic phenomena is reduced to the epistemic contingency in the microphysical phenomena: there is no further epistemic contingency in the connection.

When a phenomenon is entailed a priori by *PQTI* rather than just by *P*, something similar applies. To be sure, there may be some further epistemic contingency in *T*, *I*, and possibly *Q* so that these truths themselves may not be reductively explained. But we generally count reductive explanation “modulo totality, indexicality, and consciousness”—that is, reduction of the epistemic contingency of a phenomenon to the epistemic contingency in *PQTI*—to be reductive explanation enough. *T* and *I* are sufficiently minor additions to the reduction base that they do not change much; and while *Q* is a larger addition, scientists are prepared to put the explanation of the subjective aspects of a phenomenon (such as color and heat) on hold and settle for an explanation of the objective aspects. This “carving off” strategy arguably does not yield a complete reductive explanation of these phenomena; but at least we know just what we are not explaining.

So in a reductive explanation of a phenomenon such as water or life, we find that a low-level account of the physical processes involved will in principle imply and explain truths about the macroscopic structure, dynamics, behavior, and (in conjunction with *Q*) appearance of relevant systems. And our concepts of ‘water’ or ‘life’ dictate that systems with appropriate sorts of structure, dynamics, behavior, and appearance automatically qualify as water or as alive (at least if they are appropriately situated in our environment, or are relevantly related to systems in our environment). So the relevant microphysical truths (perhaps in conjunction with *Q*, *T*, and *I*) imply the existence of water or of life, and their existence is reductively

explained. The same applies to various specific features of water and of life, which can be implied and explained in similar ways.

The corresponding view on the problem of consciousness is what Chalmers (1996, 1999) has called “type-A materialism”: the view that there is an a priori entailment from *PTI* to the phenomenal truths *Q* (perhaps because phenomenal concepts are functional concepts). It is clear that *if* type-A materialism is true, then the phenomenal can be reductively explained in terms of the physical, for much the same reason as above. In this case, there will be a transparent epistemic connection from the physical to the phenomenal, and any epistemic contingency in the phenomenal truths will be reduced to epistemic contingency in physical truths. (And consequently, all those other macroscopic truths that are reductively explainable “modulo consciousness” will be reductively explainable *simpliciter*.) As above, a priori entailment supports reductive explanation.

Next consider a clear case of a failure of reductive explanation. Let us say that the property dualist is right about consciousness, and that consciousness is connected to the physical only by contingent laws, perhaps ultimately by fundamental psychophysical laws. These laws might be inferred, in principle, from psychophysical regularities in the actual world. Given the presence of these laws, we can still arguably have some sort of explanation of consciousness and its properties, in terms of physical processes and the psychophysical laws. From *PTI* plus psychophysical laws, the various specific phenomena of consciousness will be implied. But this will not be a case of *reductive* explanation, precisely because of the need for principles in our explanatory base over and above what is present in *P*. The laws themselves are not explained: they are epistemically primitive, in that they are not implied by more basic truths.²⁷ And these substantive, epistemically primitive principles play a central role in the explanation of the phenomena. So there is no transparent explanation of the phenomena in physical terms alone, and reductive explanation fails.

²⁷That is, they are not implied by truths that do not themselves involve phenomenal states. On a Humean view of laws, fundamental psychophysical laws will be implied by truths about the distribution of physical and phenomenal states; but this sort of implication is irrelevant to reductive explanation, as phenomenal states are what we are aiming to explain. On a non-Humean view of laws, fundamental psychophysical laws will be straightforwardly epistemically primitive.

Now let us consider a third sort of case, which corresponds to the way Block and Stalnaker envisage the situation with respect to consciousness, and to the way they envisage reductive explanation in general. Here, as in the case above, we observe numerous psychophysical regularities between brain states and states of consciousness. We are led to infer an underlying *identity* between various brain states and phenomenal states. These identities are not entailed a priori by *PTI*; and correspondingly, the truths about consciousness are not entailed a priori by *PTI*. But the truths about consciousness are entailed a priori by *PTI* plus the identities (at least if truths about brain states are entailed a priori by *PTI*). So we still have a sort of explanation of consciousness and its properties, in terms of physical processes and psychophysical identities.

This third case is a version of what Chalmers calls “type-B materialism,” on which physical truths necessitate phenomenal truths without entailing them a priori. We think that this sort of case cannot occur, but we will set that worry aside for the moment, and will pretend that it can occur. Assuming that it can occur, is the third case more akin to the first case or to the second case? There are different respects in which it resembles each. Ontologically, it is more akin to the first case, involving type-A materialism, since it is compatible with a materialist ontology on which the explained phenomenon is ultimately physical. But epistemically, it is more akin to the second case, involving property dualism.

Like the second case, the third case yields no transparent explanation of consciousness in terms of physical processes. At best, there is an explanation in terms of physical processes *plus* psychophysical identities. And epistemically, the psychophysical identities play exactly the same role as psychophysical laws. They are inferred from regularities between brain processes and consciousness, in order to systematize and explain those regularities. And most importantly, the identities are not themselves explained, but are epistemically primitive. As with the second case, it is precisely because we need these epistemically primitive psychophysical principles to explain the phenomenon that transparent reductive explanation fails.

Ontologically, these identities may differ from laws. But epistemically, they are just like laws. They are epistemically primitive psychophysical “bridging” principles that are not themselves explained, but that combine with physical truths to explain phenomenal truths. An explanation of the phenomenal will have two epistemically irreducible components: a physical component and a psychophysical component.

By calling the bridging principles identities rather than laws, this view may preserve the ontological structure of materialism. But the explanatory structure of this materialist view is just like the explanatory structure of property dualism.

Identities play a role in reductive explanation in other domains, of course. Identities involving heat, or temperature, or genes all have some explanatory work to do. But in these cases, the identities are not epistemically primitive. The identities between heat and the motion of molecules and the identity between genes and DNA, for example, are themselves implied by *PQTI* (at least insofar as they are true), for the reasons canvassed earlier in this paper. So in these cases, both macroscopic truths about the phenomena and the identities themselves are transparently explained.

It is sometimes held that “identities do not need to be explained” (for example, Papineau 1993). Block and Stalnaker say something similar (“Identities don’t have explanations”). But this seems to conflate ontological and epistemological matters. Identities are ontologically primitive, but they are not epistemically primitive. Identities are typically implied by underlying truths that do not involve identities. The identity between genes and DNA, or between water and H_2O , is implied by the underlying truths in *PQTI*, for example. Once a subject knows all the truths about DNA and its role in reproduction and development, for example, the subject will be in a position to deduce that genes are DNA. So this identity is not epistemically primitive.

Of course, just as with other truths involving macroscopic phenomena, subjects do not typically come to know these identities by deducing them from microscopic truths. But the identities are so deducible all the same, and their deducibility is what makes the phenomena in question reductively explainable. If the identities in question were epistemically primitive, then explanations of the macroscopic phenomena in terms of microscopic phenomena would have a primitive “vertical” element, and science would have established a far weaker explanatory connection between the microscopic and the macroscopic than it actually has.²⁸

Something similar applies even to the case of “Mark Twain is

²⁸One could plausibly go further, and argue that a scientist’s warrant to accept a micro-macro identity in the first place depends on a warrant to accept in-principle deducibility of the macroscopic phenomena from underlying processes. In the absence of a warrant to accept such deducibility, scientists will only be warranted in accepting correlation, not identity.

Samuel Clemens,” which Block and Stalnaker discuss. A subject who knows all the qualitative truths in question—physical, mental, social—and who possesses the concepts ‘Mark Twain’ and ‘Samuel Clemens’ will be in a position to deduce that the identity is true, even if the subject is initially ignorant of it. The subject will be in a position to know that there was an individual who was known to his parents as ‘Samuel Clemens’, who wrote books such as *Huckleberry Finn* and the like under the name ‘Mark Twain’, whose deeds were causally responsible for the current discussion involving ‘Mark Twain’ and involving ‘Samuel Clemens’, and so on. From all this information, the subject will be able easily to deduce that Mark Twain was Samuel Clemens, and the deduction will be *a priori* in the sense that it will not rely on any empirical information outside the information specified in the base. So this identity is not epistemically primitive.

We might imagine that there are eliminativists who deny the existence of Mark Twain. They accept the existence of Samuel Clemens, and the description of his exploits: that he went under the name ‘Mark Twain’, that he published many of the works that we know of under that name, and so on. But they deny that Clemens was Twain: Clemens used the name, but Twain never really existed. The response to such an eliminativist seems clear. Once the eliminativist concedes the full description of Clemens’s exploits and of his connection to our current use of the name ‘Mark Twain’, we should say: if you grant all that, you grant that Clemens was Twain. Once the qualitative description is given, there is no further fact about Twain’s existence of which we are ignorant. There is at most a terminological decision to make, and the terminological decision seems clear. So the identity of Twain with Clemens is itself deducible from underlying truths.

It may be that there is a way in which it sounds odd to ask for an explanation of why Mark Twain is Samuel Clemens, or of why water is H_2O . This is partly because we naturally interpret these phrases as asking for a causal or historical explanation. When asked why Mark Twain lived in Missouri, we will give a causal explanation of the factors that led him to be there and to stay there. It is not so easy to explain causally why H_2O is water, or why Samuel Clemens is Mark Twain (although even here, historical explanation might do at least some work).

Setting aside causal/historical explanation, it seems no harder and no easier to explain *reductively* the identities that Mark Twain is Samuel Clemens or that water is H_2O than it is to explain reductively the non-identities that Mark Twain lived in Missouri or that there is

water in the Pacific Ocean. The slight sense of linguistic oddness is equally present in all these cases, so there is nothing special about identities here. What matters is that in all these cases, the relevant truths are not epistemically primitive: they are implied by various underlying truths, and none of them are needed to deduce truths that cannot be deduced without them. So there is nothing here that gives support to the sort of epistemically primitive micro-macro identities that Block and Stalnaker need.

We have argued that (1) there is no reason to believe that there are epistemically primitive micro-macro identities, and (2) even if there were any epistemically primitive micro-macro identities, they would not support transparent reductive explanation of the macroscopic phenomena, of the sort found in most cases of reductive explanation.

Of course an opponent could hold that such identities (if they existed) would support “reductive explanation” in *some* sense—for example, a sense wherein a reductive explanation is simply an explanation that involves an ontological reduction. Nothing rests on the terminological issue, however. The significant points are (1) that this sort of explanation would be very much unlike most paradigmatic cases of reductive explanation, and (2) that by invoking an epistemically primitive bridging structure, the explanatory structure of this theory would be more akin to that of property dualism than to that of standard materialism. We might say that this sort of theory would be ontologically reductive, but epistemically nonreductive. With this sort of theory, as with a property dualist theory, the explanatory gap between physical and phenomenal might be bridged, but it would not be closed.

7. Does Materialism Require A Priori Entailment?

We have argued elsewhere that if there are phenomenal truths that are not implied by *PTI*, then materialism is false.²⁹ Block and Stalnaker deny this, endorsing a version of type-B materialism, with a conceptual gap between the physical and the phenomenal but no ontological gap. They spend much less time on materialism than they do on reductive explanation, however. Their main counterargument again proceeds via the claim that there are no such a priori entailments for most macroscopic phenomena (which are nevertheless clearly physical). We have already addressed these arguments.

²⁹Chalmers 1996, Jackson 1994, Jackson 1998.

We will not discuss the tenability of type-B materialist views in depth, as there is a large recent literature on the topic.³⁰ But it may be worthwhile to use what has gone already to outline one way of seeing what we think is wrong with the view.

The standard form of type-B materialism invokes identities to cross the explanatory gap between phenomenal states and physical or functional states. The identities are epistemically primitive, for reasons discussed above. Of course if we are right about what has gone before, this makes the physical-phenomenal identity unlike any other micro-macro identity, in which there is no analogous epistemic gap. And indeed, it makes the physical-phenomenal identity unlike any other identities, none of which are epistemically primitive in the same way.

A type-B materialist might bite the bullet on these things and hold that psychophysical identities are *sui generis*. In response, one can argue that identities between natural phenomena *cannot* be epistemically primitive. The point where one finds objective (non-indexical) epistemically primitive regularities among natural phenomena is precisely the point at which one finds fundamental natural laws. And one can argue that what it is to be a fundamental law of nature is precisely to be an objective, epistemically primitive counterfactual-supporting regularity. If this is right, then if there are epistemically primitive psychophysical regularities, they must be regarded as fundamental natural laws.

A more general version of type-B materialism holds that the physical truths *necessitate* the phenomenal truths, whether or not there are any psychophysical identities. One can argue against this view in a similar way. The view requires that there are epistemically primitive necessities connecting natural phenomena; but one can argue that all necessities connecting natural phenomena are implied by *PQTI*. Even Kripkean a posteriori necessities can be seen to be derivable by a priori reasoning from nonmodal empirical truths. So these necessities will be unsupported by analogies with any other clear cases of necessities.

Again, an opponent can bite the bullet and hold that these necessities are *sui generis*.³¹ And again, one can respond by arguing

³⁰Balog 1999, Chalmers 1999, Hill and McLaughlin 1999, Loar 1997, Levine 1998, Perry 2000, Yablo 2000, and many others.

³¹Among type-B materialists, Balog 1999 and Loar 1997 (and possibly others) accept that psychophysical necessities may be *sui generis* in this way. Balog 1999 argues that reasoning from an epistemic gap to an ontological gap is invalid, since a zombie (a physical duplicate of a conscious being, without

that necessities cannot be epistemically primitive in this way, and that the only epistemically primitive necessities connecting natural phenomena are natural (or nomological) necessities.

A more detailed treatment of the arguments here requires invoking the two-dimensional semantic framework. We will not go into those details here.³² Note that it is only at this point that the two-dimensional framework enters our arguments in an essential way, and not earlier, as Block and Stalnaker suggest.

8. Conclusions

We have argued that ordinary truths about macroscopic natural phenomena are entailed a priori by the combination of physical truths, phenomenal truths, indexical truths, and a that's-all statement. We have argued that reductive explanation requires a priori entailment, and (more briefly) that physicalism requires a priori entailment. If that is right, then if the phenomenal is reductively explainable in terms of the physical, then there is an a priori entailment from physical truths, indexical truths, and a that's-all statement to phenomenal truths. Contrapositively, if there is no such entailment, then the phenomenal is not reductively explainable in terms of the physical.

University of Arizona (Chalmers)

Australian National University (Jackson)

consciousness) could engage in similar reasoning with true premises and a false conclusion. Balog's argument crucially requires the premise that a zombie's utterance of 'I am phenomenally conscious' is true. But there are good reasons to hold that such an utterance would be false or without truth value. (In an argument between a zombie realist and a zombie eliminativist, the eliminativist would be closer to the truth.) This conclusion is supported on independent grounds in Chalmers (forthcoming b).

³²Very briefly: When the two-dimensional framework is used to argue against type-B materialism, the crucial thesis is that every a posteriori necessity has a metaphysically contingent primary intension. A posteriori necessities of the Kripkean variety all appear to have a metaphysically contingent primary intension, and are therefore compatible with the thesis. But it can straightforwardly be shown that if the thesis is true, type-B materialism is false. The only way out for the type-B materialist is to deny the thesis, accepting "strong necessities" that do not have a metaphysically contingent primary intension. The upshot is that on this view, (1) psychophysical necessities are sui generis, and that (2) one is committed to an underlying very strong modal dualism, with two distinct modal primitives (corresponding to spaces of conceptually possible and metaphysically possible worlds). Chalmers (1999) argues that this modal dualism is unacceptable.

References

- Austin, D. F. 1990. *What's the Meaning of "This"?* Ithaca: Cornell University Press.
- Balog, K. 1999. "Conceivability, Possibility, and the Mind-Body Problem." *Philosophical Review* 108:497–528.
- Block, N., and R. Stalnaker. 1999. "Conceptual Analysis, Dualism, and the Explanatory Gap." *Philosophical Review* 108:1–46.
- Byrne, A. 1999. "Cosmic Hermeneutics." *Philosophical Perspectives* 13.
- Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- . 1999. "Materialism and the metaphysics of modality." *Philosophy and Phenomenological Research* 59:473–96. <consc.net/papers/modality.html>.
- . 2002a. "The Content and Epistemology of Phenomenal Belief." In *Consciousness: New Philosophical Essays*, ed. Q. Smith and A. Jokic. Oxford: Oxford University Press. <consc.net/papers/belief.html>.
- . 2002b. "Does Conceivability Entail Possibility?" In *Conceivability and Possibility*, ed. T. Gendler and J. Hawthorne. Oxford: Oxford University Press. <consc.net/papers/conceivability.html>.
- . Forthcoming a. "The Foundations of Two-Dimensional Semantics." In *Two-Dimensionalism*, ed. M. García-Carpintero and J. Macia. *European Review of Philosophy*. <consc.net/papers/foundations.html>.
- . Forthcoming b. "The Nature of Epistemic Space." <consc.net/papers/espace.html>.
- Hill, C. S., and B. P. McLaughlin. 1999. "There Are Fewer Things in Reality than Are Dreamt of in Chalmers' Philosophy." *Philosophy and Phenomenological Research* 59:445–54.
- Horgan, T. 1984. "Supervenience and Cosmic Hermeneutics." *Southern Journal of Philosophy*, supp. 22:19–38.
- Jackson, F. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly* 32:127–36.
- . 1994. "Armchair Metaphysics." In *Philosophy in Mind*, ed. J. O'Leary-Hawthorne and M. Michael. Dordrecht: Kluwer.
- . 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
- Jackson, F., P. Pettit, and M. Smith. 2000. "Ethical Particularism and Patterns." In *Moral Particularism*, ed. B. Hooker and M. Little. Oxford: Oxford University Press.
- Nagel, T. 1974. "What Is It Like to Be a Bat?" *Philosophical Review* 83:435–50.

- Levine, J. 1993. "On Leaving Out What It's Like." In *Consciousness: Psychological and Philosophical Essays*, ed. M. Davies and G. Humphreys. Oxford: Blackwell.
- . 1998. "Conceivability and the Metaphysics of Mind." *Noûs* 32:449–80.
- Lewis, D. 1994. "Reduction of Mind." In *A Companion to the Philosophy of Mind*, ed. S. Guttenplan. Oxford: Blackwell.
- Loar, B. 1997. "Phenomenal States (second version)." In *The Nature of Consciousness: Philosophical Debates*, ed. N. Block, O. Flanagan, and G. Güzeldere. Cambridge: MIT Press.
- Papineau, D. 1993. "Physicalism, Consciousness, and the Antipathetic Fallacy." *Australasian Journal of Philosophy* 71:169–83.
- Salmon, N. 1986. *Frege's Puzzle*. Cambridge: MIT Press.
- Searle, J. R. 1991. *The Rediscovery of the Mind*. Cambridge: MIT Press.
- Yablo, S. 2000. "Textbook Kripkeanism and the Open Texture of Concepts." *Pacific Philosophical Quarterly* 81:98–122.
- . 2002. "Coulda, Woulda, Shoulda." In *Conceivability and Possibility*, ed. T. Gendler and J. Hawthorne. Oxford: Oxford University Press.
<www.mit.edu/~yablo/coulda.pdf>.