Modality and the Mind-Body Problem: Reply to Goff and Papineau, Lee, and Levine

David J. Chalmers

I am grateful to Philip Goff and David Papineau, Geoff Lee, and Joe Levine for their commentaries on *The Character of Consciousness*. All of the articles in this symposium focus on the first half of the book, especially section III on the metaphysics of consciousness. Each of them addresses my arguments against materialism and offers interesting responses to them. The main focus is on the two-dimensional modal argument in chapter 6 and on related considerations about necessity and apriority.

For background, recall that anti-materialist arguments such as the conceivability and knowledge arguments start from an epistemic gap between the microphysical truths *P* and a phenomenal truth *Q*, and from there infer an ontological gap between them. For example one can first argue that the material conditional 'If *P*, then *Q*' is not a priori, and from there argue that it is not necessary, so that materialism is false. In response, type-A materialists deny that there is an epistemic gap, typically holding that there is an a priori entailment from *P* to *Q*. By contrast, type-B materialists typically allow that there is an epistemic gap but deny that there is an ontological gap, typically holding that 'If *P*, then *Q*' is necessary even though it is not a priori.

Type-B materialists typically take inspiration from Kripke's argument that truths such as 'Water is H_2O' and 'Hesperus is Phosphorus' are necessary but not a priori. If one can make the case that 'If *P*, then *Q*' is also an a posteriori necessity, then this will strike at the heart of the anti-materialist arguments above. In response, the two-dimensional argument in effect makes a case that the materialist needs a much stronger sort of a posteriori necessity than one finds in the Kripke cases, a sort that there is good reason to reject.

According to a two-dimensional analysis, the standard Kripkean a posteriori necessities are *weak* a posteriori necessities in that their primary intensions are false at some centered metaphysically possible world: for example, the primary intension of 'water is H_2O' is false at a Twin Earth centered world (at which the watery stuff around the center is XYZ). Less technically, we can say that if we accept that the Twin Earth world is actual, we should reject ‘water is H_2O’. Likewise, the primary intension of ‘Hesperus is Phosphorus’ is false at a world where the morning and evening stars are distinct: if we accept that such a world is actual, we should reject ‘Hesperus is Phosphorus’. In these cases, the secondary intension of the key sentence is necessary, but its primary intension is contingent (equivalently, the sentence is 2-necessary but 1-contingent).

However, it is not hard to argue that weak a posteriori necessities do not help the type-B materialist. If ‘If P, then Q’ is a weak aposteriori necessity, then there will be a centered world at which its primary intension is false. With the aid of some two-dimensional analysis of the terms in *P* and *Q*, one can argue from here (setting aside a loophole for Russellian monism) that the secondary intension of ‘If *P* then *Q*’ is false. If so, the conditional is not necessary after all, and type-B materialism fails.

It follows that the type-B materialist needs to appeal to strong a posteriori necessities (or strong necessities for short): a posteriori necessities whose primary intension is true at all centered metaphysically possible worlds. To get an intuitive grip on the notion, note that any a posteriori necessity will be false at some epistemically possible scenario (more or less by definition of epistemically possible scenario). For weak necessities these scenarios will correspond to centered metaphysically possible worlds such as the Twin Earth world. For strong necessities they will not. For ‘If *P* then *Q*’ to avoid the problems above, it must be a strong necessity: the epistemically possible zombie scenario that falsifies $P\&\neg Q$ must not correspond to any metaphysically possible world where the primary intension of $P\&\neg Q$ is false.

Much of chapter 6 of *TCC* is devoted to arguing that there are no strong a posteriori necessities. I argue (i) that there are no antecedently plausible instances of strong necessities, (ii) that there is no good explanation for their existence, and (iii) that there are constitutive links between epistemic and metaphysical possibility that rule out strong necessities.

All of the commentators in this symposium in effect offer considerations in favor of strong necessities. Goff offers a semantic explanation of how there could be strong necessities. Lee argues that certain metaphysical principles about the unity of consciousness could be strong necessities. Levine argues that strong necessities need not be brute necessities and suggests that they can be explained in terms of semantically primitive concepts. Papineau suggests that epistemic and metaphysical modality have nothing to do with one another and that strong necessities are ubiquitous. In what follows, I address each of these commentators in turn.

1 Reply to Goff and Papineau

In their branching commentary, Goff and Papineau defend strong necessities in related but distinct ways. In response to my point (i) above, they jointly suggest that other a posteriori necessities such as ‘Cicero is Tully’ might be strong necessities; but they do not argue hard here and they say nothing to respond to the arguments against this claim given in the book. Instead, they in effect respond to (ii) by offering an explanation of how there could be strong necessities in principle, whether or not there are antecedendently plausible instances. They both hold that strong necessities can be generated by ‘radically opaque’ expressions—roughly, expressions with no associated descriptive content such that we have no a priori grasp of the referent. Goff provides an explanation that preserves a weaker version of the constitutive links in (iii), while Papineau argues that any such links are misconceived. I will consider these responses separately.

For present purposes, we can see my denial of strong necessities is captured by the conceivability-possibility principle CP:

(CP) For all S, S is a priori iff S is 1-necessary.

Here, S is 1-necessary if its primary intension is true at all centered metaphysically worlds. This principle is near-enough equivalent to principle CP- in chapter 6 of *TCC*: when S is ideally negatively conceivable (i.e. $\neg S$ is not a priori), S is 1-possible. In effect, for every epistemically possible scenario, there is a corresponding centered metaphysically possible world.

Goff endorses a weaker link between epistemic and metaphysical possibility. He in effect suggests an attenuated thesis TCP:

(TCP) When S is transparent, S is a priori iff S is necessary.

Here necessity is the ordinary sort of necessity, corresponding in the two-dimensional framework to 2-necessity, which requires a secondary intension true at all metaphysically possible worlds. A sentence is transparent when it involves only transparent expressions, and a transparent expression is one that expresses a transparent concept. According to Goff, a transparent concept is one that reveals the essence of the object or property that it refers to, in such a way that there are no transparent a posteriori identities. When a and b are transparent, then if $a = b$ is true it is a priori, and the same goes for $a \neq b$. Principle TCP makes the somewhat stronger claim that there are no transparent a posteriori necessities (and no transparent a priori contingencies).

Making this notion of transparency precise involves a number of tricky issues, but I think it corresponds at least roughly to a notion familiar from the two-dimensional framework. We can say that an expression is transparent iff it is super-rigid: that is, it has a rigid two-dimensional intension, picking out the same referent at every epistemically possible scenario and every metaphysically possible world (and at every pair thereof). All super-rigid expressions are epistemically rigid, in that they pick out the same referent at every epistemically possible scenario. Intuitively, an expression is epistemically rigid iff one can know a priori what it refers to. Expressions such as 'water' and 'Gödel' are not epistemically rigid, super-rigid, or transparent, but expressions such as 'zero' are plausibly epistemically rigid, super-rigid, and transparent.

With this background, it is not hard to see that CP entails TCP. The primary and secondary expressions of a super-rigid expression pick out the same referent at a world (centering aside), so the primary and secondary intensions of a transparent sentence have the same truth-value at a world. So a transparent sentence is 1-necessary iff it is 2-necessary, and the entailment follows. In reverse, TCP does not obviously entail CP, as TCP makes no prediction about what happens in cases not involving transparent expressions. In effect, TCP is a close relative of some theses discussed on pp. 203-205 of the book (involving notions of semantic stability, semantic neutrality, and cognitive transparency), all of which are weaker than CP in a similar way.

Goff then has two key points. First, a materialist who holds that the concept of phenomenal consciousness is radically opaque can deny CP without denying TCP. Second, the weaker thesis TCP suffices to retain a reasonable modal rationalism with constitutive links between the epistemic and modal domains and without a primitive metaphysical modality. If this is right, then a reasonable modal rationalism is compatible with a type-B materialist view of consciousness.

This is undeniably a challenging and interesting position. There is much to say about the second point and whether TCP without CP can satisfy the motivations for modal rationalism. But for reasons of space here I will concentrate on the first point, concerning radical opacity.

Goff and Papineau say that a term or concept is radically opaque iff it does not reveal any substantive information about its referent. This characterization is itself somewhat opaque, but if we take it that the relevant "revealing" must be a priori and that the substantive information must be presented transparently, then we can say that an expression C is radically opaque iff there is no transparent predicate ϕ such that '$\phi(C)$' is a priori. An obvious worry here is that candidates for ϕ such as 'is an object', 'is a property', 'is self-identical' suggest that no expression will be radically opaque. If one avoids the worry by requiring that the transparent predicate be an *identifying predicate* (e.g. one of the form 'is the F'), then the worry is that even on my view ordinary proper names such as 'Gödel' will be radically opaque, as they are not a priori equivalent to transparent descriptions.

Still, within my own framework I think we can understand radical opacity in terms of speakers' ability to identify a term's referent within a scenario described in transparent terms. Following the framework in *Constructing the World*, we can say that a sentence S is *scrutable* from a sentence T when the material conditional 'If T, then S' is a priori. Then we might say that C is radically opaque when no true sentence of the form 'C is the F' is scrutable from a sentence T where both F and T use only canonical vocabulary. We might restrict canonical vocabulary to transparent expressions, perhaps along with primitive indexical expressions such as 'I' and 'now'.

If there were radically opaque expressions in this sense, it would cause problems for the two-dimensional argument. A sentence's primary intension is in effect defined in terms of scrutability from scenarios specified in canonical vocabulary, and at least where primary intensions over centered metaphysically possible worlds are concerned, canonical vocabulary should include only transparent expressions and primitive indexical expressions. So if there are radically opaque expressions, their intensions will not be well-defined and the two-dimensional argument will break down. (Intensions over epistemically possible scenarios will still be well-defined, but these worlds need not correspond to centered possible worlds.)

I think there are no radically opaque expressions in this sense (with a qualification discussed shortly). In chapter 7 of *TCC*, I argue that all ordinary macroscopic truths are scrutable from $PQTI$, which can serve as a canonical vocabulary. In *Constructing the World*, I argue at length that all truths are scrutable from $PQTI$. If this is right (and neither Goff nor Papineau provide any reasons to reject these arguments), it eliminates at least a certain sort of radical opacity: opacity of the nonbase vocabulary relative to the base vocabulary.

Still, the suggestion that phenomenal concepts are radically opaque suggests another sort of opacity: opacity in the base vocabulary itself. If Q involves nontransparent and nonindexical expressions, and is not itself scrutable from a sentence involving transparent and indexical expressions, then there will be no base vocabulary involving only these expressions, and the argument will still be in trouble.

On my view, all opacity arises either from primitive indexicals, or from nonbasic concepts that are scrutable from basic concepts. Non-indexical basic concepts themselves are transparent. There is one way one can make this conclusion trivial: by counting any non-transparent basic concept as a primitive indexical concept. The primitive indexicals needed in the base already includes not just 'I' and 'now' but certain demonstratives such as 'this$_E$' as a demonstrative for an experience.

If 'consciousness' itself were nontransparent, we might see it as a sort of demonstrative: "that property", with some sort of ostension of a property that we do not fully grasp.

Of course this just moves the bump in the rug: if core phenomenal concepts are demonstrative concepts, then we will need elements in the center of a centered world corresponding to their referents, and it will be unclear why these elements cannot be physical properties. Still, there is good reason to believe that our core phenomenal concepts are not demonstrative concepts. I argue this at length in chapter 8 of *TCC*. We certainly have demonstrative concepts of phenomenal properties ("that property"), but our central phenomenal concepts are distinct from these, as brought out by the cognitive significance of identities linking them to demonstrative concepts: for example, Mary might have cognitively significant knowledge of 'that property is phenomenal redness'. This suggests that these core concepts are not demonstrative or indexical concepts, and indeed suggests that we have some substantive grasp of their referents, so that they are not radically opaque.

Perhaps someone will suggest that there are different sorts of phenomenal demonstratives here, or that phenomenal concepts can be radically opaque even though they are not demonstrative. In chapters 6 and 10 of TCC, I argue against some models of phenomenal concepts that might be construed in the latter sort, but until a model is fleshed out in some detail it is hard to argue against it.

In any case, I think there are good reasons to believe that phenomenal concepts are epistemically rigid. That is, anyone who possesses a phenomenal concept is in a position to know a priori what phenomenal property (a certain shade of red34, say) it picks out. These concepts are also plausibly super-rigid: as I argue in chapter 7 of *TCC*, a pure phenomenal concept picks out the same phenomenal property across all epistemically possible scenarios and metaphysically possible worlds. Goff himself has argued in other work that phenomenal concepts are transparent, and I am inclined to endorse his arguments. If phenomenal concepts are transparent (or if they are super-rigid), the objection will not arise.

Moving beyond phenomenal concepts, I think there are good general reasons to believe that our most fundamental concepts are either transparent concepts or indexical concepts, and that all truths are scrutable from a base of truths involving only these concepts. I make this case in Excursus 14 of *Constructing the World*. If this is right, then no concepts will be radically opaque in Goff's sense. If so, then Goff's reasons for rejecting CP will fail, and his reasons for accepting TCP will also provide reasons for accepting CP.

Papineau's view is that my modal rationalism should be rejected on the grounds that epistemic and metaphysical modality have nothing to do with each other. He brings this out by arguing that facts about metaphysical modality cannot be grounded in facts about conceivability. For example, ‘David Papineau’s father is Owen Papineau’ is metaphysically necessary but it is not epistemically necessary, and one cannot explain its metaphysical necessity in terms of facts about epistemic necessity.

My view is not quite the view that metaphysical modality can be explained entirely in terms of epistemic modality. Rather, I think that metaphysical modality can be explained in terms of epistemic modality plus conceptual structure, along with nonmodal truths about the actual world. The relevant conceptual structure is two-dimensional conceptual structure, intuitively capturing a priori truths about the way that expressions apply to epistemic possibilities and the way they apply to counterfactual possibilities depending on which epistemic possibility turns out to be actual.

The framework invokes a single space of worlds, characterized in terms of epistemic modality. One way to develop the proposal, suggested in *TCC* and also sharing the spirit of Goff’s proposal using TCP, is to define worlds in terms of (equivalence classes of) of maximally specific epistemically consistent transparent sentences. One can define centered worlds by conjoining these transparent sentences with certain indexical sentences. One can then associate an expression with a primary intension over centered worlds: a sentence S will be true at a centered world w specified by D iff $D \rightarrow S$ is a priori. One can also associate expressions with a two-dimensional intension over ordered pairs of centered worlds and worlds: S is true at such a pair (w, w') specified by D and D' iff $D \rightarrow \Box(D' \rightarrow S)$ is a priori. Then D is metaphysically necessary iff its two-dimensional intension is true at $(@, w)$ for all worlds w, where $@$ is the actual world.

This is not a reductive definition of metaphysical necessity in terms of epistemic necessity alone. If we see the definitions of intensions as part of the account of metaphysical necessity, then the definition of a two-dimensional intension itself invokes the notion of metaphysical necesity. And if we take two-dimensional intensions as basic here, it is clear that the second dimension nonetheless captures something about the way an expression applies to counterfactual possibilities. But the key point is that here we are invoking only a priori truths involving metaphysical necessity, and indeed plausibly only conceptual truths involving metaphysical necessity, so that the two-dimensional intension can be seen as a sort of conceptual structure.

For example, ‘Necessarily, Hesperus is Phosporus’ is an a posteriori truth. But ‘If Hesperus is Phosphorus, necessarily Hesperus is Phosphorus’ is plausibly a priori and indeed a conceptual truth, as it is plausibly part of the conceptual structure associated with ‘Hesperus’ that it is a rigid designator. In effect, this structure is built into the two-dimensional intension associated with ‘Hesperus’. So the a posteriori metaphysical necessity here can plausibly be explained in terms of conceptual structure along with the nonmodal truth that Hesperus is Phosphorus in the actual world. Likewise, it is plausible a priori that 'If the underlying explanatory structure of water is H_2O, then necessarily water is H_2O'. It is plausibly part of the conceptual structure associated with 'water' that it applies counterfactually to whatever shares the relevant underlying explanatory structure, and in effect this structure if built into the two-dimensional intension.

The current model generalizes this behavior to all metaphysical necessities. It in effect invokes the following general principle (MNM) (for "modal/nonmodal").

> (MNM) For every a posteriori metaphysical necessity 'Necessarily *S*', there is a nonmodal truth *D* such that 'If *D*, then necessarily *S*' is a priori.

This principle is certainly not undeniable: some who hold that there are strong necessities such as 'There is an omniscient being' may deny it. Still, it is a plausible principle that seems consistent with Kripke's core cases of the necessary a posteriori while also being congenial with a modal rationalist framework. One we combine this with the thesis that every nonmodal truth is scrutable from truths in transparent and indexical vocabulary, one obtains a stronger version of (MNM) in which *D* is restricted to that vocabulary, thereby cohering with the definition of two-dimensional intensions given above. Furthermore, it is plausible that the a priori conditionals in question are grounded in conceptual structure: the two-dimensional conceptual structure of the expressions in *S*. In this way the metaphysical modality is grounded in conceptual structure, nonmodal truths, and a priori truths. So there is no need to make unreduced appeal to primitive substantive facts about the metaphysical modality.

With this much in hand, it is easy to see how the framework handles Papineau's core case. Take the sentence 'OP is DP's father', and let us go along with the Kripkean view that this sentence is metaphysically necessary. It is then plausible that 'If OP is DP's father, then necessarily OP is DP's father' is a priori, fitting the template of (MNM) above. Furthermore, it is plausible that this a priori conditional reflects underlying conceptual structure: roughly, conceptual structure determining that insofar as 'DP' refers to a person with parents, it refers to a person with those parents in all possible worlds. In a little more detail: 'DP is a person' is either a conceptual truth or is scrutable from underlying nonmodal truths, and it is also plausibly a conceptual truth that 'If DP is a person, DP is necessarily that person'. If one accepts the Kripkean framework, it is highly plausible that 'If a person has certain parents, that person necessarily has those parents' is a priori. We can see this as part of the two-dimensional conceptual structure associated with 'person'. Derivatively, 'DP' will have conceptual structure such that, given a scenario in which

‘DP’ refers to a person with parents, ‘DP’ will pick out a person with those parents in all possible worlds.

By contrast, if we consider ‘DP’s birthplace is Como’, the following is clearly not a priori: ‘If a person has a certain birthplace, that person necessarily has that birthplace’. So constancy of birthplace is not built into the conceptual structure of ‘person’ or ‘DP’. It is entirely possible that there is a language in which there is a term ‘schmerson’ such that ‘If a schmerson has a certain birthplace, that schmerson necessarily has that birthplace’ is a priori, and on which there is a related term ‘DP*’ such that ‘DP* is a schmerson’ is true in that language. That language is not English. But insofar as there is parity between the two languages, this brings out all the better that the necessity of origins turns on features of our concepts rather than substantive metaphysical truths about the world.

Papineau’s article suggests an objection: even if this account appeals to a priori conceptual truths, they are still a priori conceptual truths about metaphysical necessity, so the whole account presupposes the notion of metaphysical modality and does not explain it. At this point I am happy to allow that there are distinct notions of epistemic modality and metaphysical modality as applied to sentences, neither reducible to the other. These two notions are grounded in the epistemic and subjunctive evaluations of expressions in possibilities, as one finds in indicative and subjunctive conditionals. Neither of these kinds of evaluation is reducible to the other, and in that sense the account takes subjunctive evaluation (a metaphysically modal notion) as primitive. However, both kinds of evaluation are sorts of semantic evaluation that reflect underlying conceptual structure. And crucially, the two dimensions of evaluations do not require two distinct spaces of possibilities (centering aside) at which expressions are evaluated. There is just one underlying space of possible worlds, one that need not be characterized in terms of the metaphysical modality at all. This is all that the modal rationalist picture that I favor claims.

(For discussion of some further issues concerning de re modality, see footnote 3 on pp. 188-89 of *TCC*.)

2 Reply to Lee

Geoff Lee focuses on my discussion (co-authored with Tim Bayne) of the unity of consciousness in chapter 14 of *TCC* and uses it to raise difficulties for the modal rationalism outlined in chapter 6. He focuses on the relationship between total states of consciousness (what it is like to be subject at a time) and local or atomic states of consciousness (what it is like to experience a certain color at a location, perhaps). I hold that the total state subsumes the local state, where subsumption is a quasi-mereological relation. But this claim is compatible with different theses about the priority between the two.

Lee distinguishes three different theses: holism, according to which the total states are primitive and atomic states derive from them; atomism, according to which the atomic states are primitive and total states derive from them; and a no-priority view, according to which neither sort of state derives from the other. He argues in effect that one of these three theses is a strong necessity. We might summarize his argument as follows:

(1) Either atomism, holism, or the no-priority view is true.

(2) If one of these three theses is true, it is necessarily true.

(3) None of these three theses is a priori.

(4) These three theses involve only super-rigid expressions.

(5) Any necessarily true thesis involving only super-rigid expressions that is not a priori is a strong necessity.

---

(6) One of atomism, holism, or the no-priority view is a strong necessity.

Lee uses this argument to suggest that consciousness has a hidden essence of a sort that I am committed to denying.

The issue of holism versus atomism in the structure of consciousness is directly analogous to the issue of monism vs. pluralism (Schaffer 2010) in the structure of the world. The (priority) monist holds that the world and its states are primitive, and that the parts and their properties derive from this, while the (priority) pluralist holds that the priority is in the other direction. Indeed, someone might mount an argument analogous to Lee's to make the case that monism, pluralism, or a related thesis is a strong necessity. These arguments will be of a piece with other arguments for strong necessities in metaphysics, arguing that various substantive metaphysical theses (mereological universalism, say) are necessary if true without being a priori either way. But I will concentrate on Lee's arguments about consciousness here.

In response to Lee's argument I am inclined to deny (3). In the chapter, Bayne and I tentatively favor holism, on the grounds that our basic concept of consciousness is the concept of a total conscious state (what it is like to be a subject at a time). These grounds might be turned into an a priori argument for holism, as follows.

(1) The concept of a total conscious state is conceptually prior to the concept of an atomic conscious state.

(2) Both sorts of concept are super-rigid.

(3) Conceptual priority relations among super-rigid concepts mirror metaphysical priority relations among their referents.

–

(4) Total conscious states are metaphysically prior to atomic conscious states.

Of course one could question these premises or their apriority. I do not think that premise (1) is obvious, but I think that it has some plausibility on reflection. And plausibly its truth or falsity is an a priori matter, as it involves only the conceptual and not the metaphysical domain. Premise (3) can certainly be denied, but I argue for it in *Constructing the World* and think an a priori case for it can be made. So I think there are at least some reasons for thinking that holism is a priori.

Lee makes a few objections to holism. He raises three initial objections to complex primitive states and answers them himself by an analogy with the quantum-mechanical wavefunction. I endorse his answers.

Lee also argues that holism is hard to reconcile with type-F or Russellian monism, on which our conscious states are constituted by phenomenal or protophenomenal states of microphysical entities. These views are obviously inconsistent with a version of holism according to which our total conscious states are primitive. Still, they are not obviously inconsistent with a version of holism on which our total conscious states are *relatively* primitive: that is, primitive relative to our atomic states. If we insist on understanding holism, atomism, and so on in terms of absolute primitiveness, then one can reasonably deny premise (1) of Lee's argument, and I am no longer sure that I am a holist. If we understand it in terms of relative primitiveness, then it is consistent with Russellian monism. Given the panpsychist version of Russellian monism, a holist will presumably hold that the total phenomenal states of microphysical entities are either prior to or identical to their atomic phenomenal states, that these microphenomenal states help to constitute our total macrophenomenal states, and that the macrophenomenal state constitutes our macrophenomenal atomic states in turn. There is no obvious incoherence here.

That said, there do remain serious questions about how to reconcile Russellian monism with holism and the unity of consciousness. It is not at all obvious how a huge cluster of microphenomenal or protophenomenal properties should combine to yield a unified total macrophenomenal state. It is much easier to see why consciousness should come in unified holistic bundles on a substance dualist view on which selves and their conscious states are both fundamental. Insofar as I divide my credence in the metaphysics of consciousness between substance dualism and Russellian monism, I am inclined to be more confident about the unity of consciousness under the first hypothesis than under the second.

Lee also considers a priori arguments for holism along the lines above and suggests that they establish only that *if* there are subjects of experience then holism is true (of the experiences of those subjects?). But Lee thinks it is not a priori that all experiences have a subject; so the arguments do not establish the apriority of holism. Now, I think that in an ordinary sense of "subject", it is very plausibly a priori that experiences must have a subject. Lee appears to have in mind a more loaded conception of "subject" involving a single self "standing behind" every component of one's consciousness. I am not sure that I fully understand this more loaded sense, so I am not sure whether it is a priori that experiences must have a subject in this sense.

Furthermore, if it is not a priori that experiences must have a subject in Lee's sense, it may well not be necessary either. Perhaps some experiences have a subject in Lee's loaded sense and some do not. If so, and if Lee is right about the connection to holism, holism will be true of those experiences that have subjects but not those that do not. On this view we will be led to reject premise (2) of Lee's argument. Maybe Lee would respond by suggesting that the thesis that all experiences have a subject is itself necessarily true if true, without being a priori either way. But now the original issue about holism is no longer doing the work in the argument for strong necessities; all the work is being done by an analogous argument about the thesis that all experiences have subjects. Here the whole dialectic will recapitulated; certainly I think it is far from clear that analogs of premises (2) and (3) are both plausible here.

For example, one especially loaded conception of a subject is one on which subjects must be metaphysically fundamental entities, as on a substance dualism views. Let us call these entities Subjects. One might at least speculate that intuitions about holism and indeed about the necessary unity of consciousness rest on the tacit premise that experiences have Subjects. I confess that I have days on which it seems to me that it is a priori that all experiences have Subjects, for example because it is a priori that all experiences have subjects and because our ordinary concept of subjects does not allow them to be grounded in more basic entities (one might suggest that for any such entities, it is conceivable and therefore possible that they exist without subjects). I also have days when this does not seem a priori. But even under the hypothesis that some experiences do not have Subjects, I think it remains conceivable and possible that some experiences have Subjects. If so, it is not part of the essence of experience that it has a Subject and not part of the essence that it does not. And given that holism and unity go along with having a Subject, it is not part of the essence of experience that it is holistic or unified and not part of the essence that it is not.

Moves of this sort, denying Lee's premise 2, can also be made more directly to the original argument. If I were to be convinced of Lee's premise 3 and held that neither holism, atomism, or their negations is a priori, I would then hold that both holism and atomism are conceivable and that both are possible. Perhaps there are two sorts of experience, such as those had by a Subject and those that are not, such that holism is true of one sort but not the other sort.

Lee considers a response of this sort and objects that it requires that the same total phenomenal state could be enjoyed in a holistic way in one world and an atomistic way in another world. I do not think this claim follows: perhaps top-down experiences differ in some systematic phenomenal way from bottom-up experiences. But even if the claim is true, I do not think it is obviously incoherent. As Lee notes, there are mass properties that are sometimes realized primitively and sometimes realized derivatively. It is not out of the question that phenomenal properties could be multiply realizable in this way. Lee suggests that in the phenomenal case, theses such as holism must reflect the nature of the properties rather than just their realization, so that they will be necessary if true. Once I accept that both holism and atomism are both conceivable where the same phenomenal property is concerned, though, I lose any strong intuition that these must be built into the nature of the property.

3 Reply to Levine

Joe Levine focuses on the A Priori Entailment thesis (AE), which holds that all truths are a priori entailed by metaphysically fundamental truths (or equivalently, that all truths are scrutable from metaphysically fundamental truths). He says that I argue for AE in two ways: by arguing that denying AE requires brute necessities, and by arguing that any apparent counterexamples to it are not counterexamples at all. He responds to the first argument by arguing that one can deny AE without embracing brute necessities. He responds to the second argument by arguing that my appeal to phenomenal truths in the a priori entailment base PQTI begs the question against the materialist.

Levine's vision of the dialectical situation differs significantly from mine. Although I am sympathetic with thesis AE, it does not play a central role in my main arguments against materialism in *TCC*. The main anti-materialist arguments (in chapters 1 and 6) do not mention AE: rather, the argument in chapter 1 rests on considerations about functional explanation, and the argument in chapter 6 rests on the two-dimensional conceivability-possibility thesis CP. Theses akin to AE play a minor role in chapter 5 and become central only in chapter 7, where Frank Jackson and I are concerned mainly with rebutting putative counterexamples to related theses raised by Ned Block and Robert Stalnaker. Furthermore, while my argument for CP turns on an argument that there are no strong necessities, the objection that strong necessities are brute necessities plays only a minor role here, and the main argument against strong necessities is quite different. The overall result is that Levine directly addresses anti-materialist arguments that I do not give or that are not central, while my central arguments, especially the two-dimensional argument in chapter 6 that proceeds via thesis CP, are not directly addressed.

Still, the arguments that Levine addresses are interesting ones, and I am not unsympathetic with them. The claim that strong necessities will be brute necessities played a more central role in *The Conscious Mind*, and I briefly use a thesis akin to AE (which I call Fundamental Scrutability) to argue against materialism in *Constructing the World*. And some aspects of Levine's objections to these arguments might also be used to object to my argument from thesis CP. So I will say something about them in what follows.

On brute necessities: In *The Conscious Mind* (pp. 136-8) I suggested that strong necessities (unlike weak necessities) will require brute and arbitrary restrictions on the space of possible worlds. These constraints will not derive from a priori reasoning (which restricts us only to conceivable worlds) or from empirical observation (which only locates us among the worlds without restricting the space), and will require something akin to primitive, unexplained modal principles. Levine responds, in effect, that the strong necessities that the type-B materialists needs need not be brute necessities. In particular, a psychophysical necessity (e.g. $\Box(m = p)$) is explained by the a priori truth that identities are necessary (if $m = p$, then $\Box(m = p)$) and a nonmodal truth, namely the identity itself ($m = p$).

In chapter 6 (p. 191) I briefly discuss a related strategy that relies on property identities to avoid problems for strong necessities, and respond by saying that properties themselves subject to modal constraints: if two properties could have come apart, they are not the same property. In effect a property identity builds in a modal claim, and the reasons for rejecting strong necessities that violate the CP principle are also reasons for rejecting "strong identities" that violate the principle. In effect, to claim that $m = p$ is not just to claim that m and p are actually coextensive but that they are necessarily coextensive. There are delicate issues of priority here, and Levine might argue that the necessity is here grounded in the identity and so is not brute. But then the worry is that

$m = p$ is a brute and inexplicable identity, which is just as problematic as a brute and inexplicable necessity.

Here we enter familiar territory about whether there can be inexplicable identities and whether identities ever need explanation. Frank Jackson and I respond to the popular line that identities do not need explanation in chapter 7 (pp. 244-5), saying that it conflates ontological and epistemological questions. Identities are ontologically primitive, but explanation is epistemological: just as knowing $a = b$ requires knowledge that connects the distinct modes of presentation in a and b, explaining $a = b$ requires an explanation that connects these modes of presentation, and this is highly nontrivial. In the case of a standard scientific identity, such a mode-of-presentation-connecting explanation can be given and explains the identity. But for $m = p$ (as Levine has famously noted) these explanations do not seem to be available. And as elsewhere in science, truly inexplicable claims can only be countenanced at the fundamental level.

Levine notes that we might have reason to accept an identity even in the absence of a mode-of-presentation-connecting explanation. That may be true: justification for believing p and explanation of p are different things. But if I am right, then such an explanation must at least be possible for the identity to be true. So if such an explanation is impossible, the identity must be rejected.

Of course this dialectical situation raises many delicate questions about explanation. The argument against strong necessities that I gave in *TCC* (that they lead to a problematic modal dualism) has the advantage of avoiding those questions, or at least replacing them with delicate issues about modality that are perhaps a little less murky. Levine does not address that argument here.

Levine's second major objection focuses entirely on the discussion in chapter 7, which is a revised version of "Conceptual Analysis and Reductive Explanation", co-authored with Frank Jackson. It is worth stressing that this chapter focuses almost entirely on epistemological issues rather than on ontological issues such as materialism. The focus is squarely on the question of whether ordinary macroscopic truths are a priori entailed by microphysical truths perhaps along with phenomenal truths, and not on the upshot for materialism. Even Levine's thesis AE about a priori entailment from fundamental truths is not mentioned (though a version of it was mentioned briefly late in the original article). The main place in the chapter where ontological issues are mentioned is in the new footnote 1, which responds to Levine 2010 (which presented a related critique of the original article) by noting explicitly that ontological issues are bracketed in this chapter! Levine acknowledges the footnote but proceeds as if ontological issues are central all the same.

Still, the anti-materialist argument that Levine discusses is an important one. The argument proceeds as follows.

(1) All truths are scrutable from the fundamental truths [AE].

(2) It is not the case that all truths are scrutable from the microphysical truths.

–

(3) The microphysical truths do not exhaust the fundamental truths.

In response, a type-B materialist will deny AE. So the nature of arguments for AE will be crucial here. One way to argue for AE is to argue that it follows from a conceivability-possibility principle such as CP; Levine does not discuss arguments of this sort. Another way to argue for AE is to argue that it fits all the familiar cases (leaving aside cases tied to consciousness). It is arguments of this sort that Levine focuses on.

Arguments from familiar cases to AE and arguments from AE to anti-materialism do not play a central role in *TCC*, but they play a subsidiary role in *The Conscious Mind* and *Constructing the World*, and the arguments given in chapter 7 of *TCC* can be used to support them. Strictly speaking the relevant thesis is not AE but the related thesis AE* holding that all truths are scrutable from fundamental truths conjoined with a totality truth *T* and an indexical truth *I*. A type-A materialist such as David Lewis might argue from familiar cases to AE* by arguing that all familiar macroscopic truths are scrutable from *PTI*, the conjunction of all microphysical truths *P* with *T* and *I*. That strategy is not available to a anti-materialist like me: phenomenal truths will certainly not be scrutable from *PTI*, and nor will truths with a strong dependence on phenomenal truths. Instead, Jackson and I proceed neutrally by arguing that all truths are scrutable from *PQTI*, which includes the conjunction *Q* of all phenomenal truths in addition. Call this thesis *PQTI*-scrutability.

Levine does not dispute *PQTI*-scrutability here. (He says that one might deny it by denying the Fregean semantics on which it is based, but the argument for *PQTI*-scrutability does not assume any particular semantic theory.) Instead he denies that it supports AE*. The most obvious direct argument from *PQTI*-scrutability to AE* requires the additional premise that the fundamental truths include *P* and *Q*. This premise will be accepted by dualists but rejected by materialists. It is for this reason that Levine says the argument begs the question against materialists. That is correct, but this direct argument is not the relevant argument for AE*.

A better argument proceeds as follows. From the initial premise of *PQTI*-scrutability, we infer that *PTI* entails all truths that are independent of consciousness. It follows that setting aside the consciousness-dependent truths (where dependence is understood epistemologically), AE* is true in all cases. This thesis is precisely what both dualists and physicalists who accept AE* would expect, so it provides significant abductive support for AE*. Of course the last step is not deductive, so type-B materialism remains consistent with the data. But the type-B materialist is left in the uncomfortable position of saying that consciousness-dependent truths are the only exceptions to AE*. Unless the materialist can give a compelling explanation of why there should be this special exception to the principle, the obvious response is to reject the exception as ad hoc and accept AE* in full generality.

Because this argument for AE* is abductive rather than a deductive argument, it works best when combined with independent arguments for AE* such as the argument from CP. But I think the argument nevertheless has force and certainly does not beg any questions. And I think it captures well the dialectical situation that we find ourselves in, in which type-B materialists have been trying hard to meet the burden of explaining why there should be special exceptions to principles such as AE and CP in the case of consciousness.

In the last part of his discussion Levine takes up this burden himself and offers such an explanation. He suggests that the existence of consciousness-dependent exceptions to AE can be explained by the hypothesis that the concept of consciousness is semantically primitive, so that it does not enter into a priori entailments with other concepts. Semantic primitiveness does not entail metaphysical primitiveness, so this way of explaining exceptions to AE* is entirely compatible with materialism.

Here in effect Levine is suggesting a version of the so-called phenomenal concept strategy: explaining the special epistemic gaps associated with consciousness in terms of special features of phenomenal concepts. Indeed, a basic commitment of that strategy is to accept conceptual dualism without ontological dualism, so the claim that the concept of consciousness is conceptually primitive is typically a key part of the strategy. Most existing versions of the strategy try to explain this claim in terms of further features of phenomenal concepts, for example by holding that they are indexical or recognitional or quotational concepts. Levine does not offer such a further explanation, but he articulates what many of these strategies share.

It is also worth noting that the same strategy can be used to respond to my arguments for CP. Here the claim will be that we should expect exceptions to CP when we have semantically primitiveness unaccompanied by metaphysical primitiveness. So the special exception of strong necessities in the case of consciousness can be explained in terms of semantic primitiveness without metaphysical primitiveness.

A few things are worth noting quickly. First, semantic primitiveness alone does not explain the relevant epistemic gaps, as there can be a priori entailments between truths invloving semantically primitive concepts (e.g. between *red* and *color*, between natural and normative truths, between any truths and mathematical truths). Second, the semantic primitiveness of phenomenal concepts itself requires explanation, and there is good reason (see chaper 10 of *TCC*) to think that no adequate account that also explains the epistemic gaps can be given in physical terms.

I am also inclined to think that there are strong connections between semantic and metaphysical primitiveness. In *Constructing the World* (excursus 15), I tentatively endorse the thesis that in the special case of truths expressing epistemically rigid concepts, then any metaphysical grounding relations among these truths are mirrored by conceptual grounding relations. Here one thought is that epistemically rigid concepts reveal the nature of their referents, so one should expect metaphysical grounding relations that spring from that nature to be conceptually transparent. With additional minor assumptions, it follows that any primitive epistemically rigid concept picks out a metaphysically primitive property. Of course there are cases where semantic primitiveness and metaphysical primitiveness come apat, including concepts such as *I* (in one direction) and *charge* (in the other), but the obvious cases all involve epistemic nonrigidity. So if we accept that consciousness is epistemically rigid, as I do, then its semantic primitiveness may not be dissociable from metaphysical primitiveness.

These waters are murky enough that I would not want to rely on the concept of semantic primitiveness or on its potential connections to metaphysical primitiveness in order to argue against materialism. But by the same token, I do not think the materialist can rely on semantic primitiveness and its putative dissociation with metaphysical primitiveness in order to resist arguments for materialism. But perhaps both sides can agree that sorting out these murky issues has the potential to shed considerable light on the mind–body problem.

References

Chalmers, D.J. 2010. *The Character of Consciousness*. Oxford University Press.

Levine, J. 2010. The Q factor: Modal rationalism versus modal autonomism. *Philosophical Review* 119: 365-80.

Schaffer, J. 2010. Monism: The priority of the whole. *Philosophical Review* 119: 31-76