Conceivability and the Metaphysics of Mind

Joseph Levine

Noûs, Vol. 32, No. 4. (Dec., 1998), pp. 449-480.

Stable URL:
http://links.jstor.org/sici?sici=0029-4624%28199812%2932%3A4%3C449%3ACATMOM%3E2.0.CO%3B2-7

Noûs is currently published by Blackwell Publishing.

NOÛS 32:4 (1998) 449–480

Conceivability and the Metaphysics of Mind

JOSEPH LEVINE
North Carolina State University

Introduction.

Materialism in the philosophy of mind is the thesis that the ultimate nature of the mind is physical; there is no sharp discontinuity in nature between the mental and the non-mental. Anti-materialists assert that, on the contrary, mental phenomena are different in kind from physical phenomena. Among the weapons in the arsenal of anti-materialists, one of the most potent has been the conceivability argument. When I conceive of the mental, it seems utterly unlike the physical. Anti-materialists insist that from this intuitive difference we can infer a genuine metaphysical difference. Materialists retort that the nature of reality, including the ultimate natures of its constituents, is a matter for discovery; an objective fact that cannot be discerned a priori.

In this paper I undertake to provide an explicit analysis of the dialectic that surrounds the conceivability argument. My principal conclusion is that the materialist is right in resisting the reasoning that starts from considerations of what is conceivable and ends with genuine metaphysical conclusions. However my approach is much more sympathetic to the anti-materialist position than is the standard materialist line, and I will provide a limited defense of some crucial aspects of the anti-materialist position. Materialism will emerge from this fight intact, but shaken.

I will proceed as follows. In section 1 I will address general questions concerning modal intuitions, the relation between conceivability and possibility, and the like. It's very difficult to avoid controversy on these topics, but I hope that what I have to say will be untendentious enough to serve as a background to my main concern, which is the applicability of conceivability considerations to the mind-body problem. In sections 2 and 3 I will present and criticize the standard anti-materialist conceivability argument, but then in section 4 I will present a stronger version of the conceivability argument, one immune to the

objections of the previous sections. In section 5 I will consider replies to this stronger argument.

1.

Let us assume that possible worlds are composed of situations, or facts, which are distributions of properties over objects. The maximal set of possible worlds is the set of metaphysically possible worlds, and when we say that some situation is possible, without qualification, we mean that there is a metaphysically possible world in which it obtains. To say that a situation is nomologically possible, by way of contrast, is to say that it obtains in a world governed by the same laws as the actual world. I assume that the set of nomologically possible worlds is a proper subset of the set of metaphysically possible worlds.

Modal intuitions, or judgments, concern what is possible and what is impossible. We judge it impossible that it should be both raining and not raining, and judge it possible that Clinton could have lost the election of 1996 and hence not been the first twice-elected Democratic President since Roosevelt. How is it that we can ascertain of the latter situation that it obtains in at least one possible world, and of the former that it doesn't? What provides us epistemic access to metaphysical possibility and necessity?

It seems to me that our cognitive access to modal facts is primarily a matter of our sensitivity to the logical forms of the representations by which we conceive them. I just don't see what else modal intuition could be, unless we allow in some quite dubious and mysterious mental powers. But this doesn't immediately solve the problem of epistemic access to modal facts. To do that we must connect the formal features of representations to the modal facts themselves, which are metaphysically independent of minds and their representations.

I propose to do this through the following Modal Bridge Principle:

(MBP) A situation S is metaphysically possible just in case it has no accurate representation that is logically inconsistent.

R is an "accurate representation" of S just in case its terms pick out the objects, properties and relations composing S, and its compositional mechanisms reflect the manner of their relation in S.[1] By "logically inconsistent" I intend a purely formal notion. I do not intend MBP as a definition of metaphysical possibility, since I don't believe that formal consistency, which applies to representations, is somehow more basic than the metaphysical possibility of a situation, which is a representation-independent notion. Rather, I think of the two notions involved—formal consistency and metaphysical possibility—as interdefinable, with neither one more basic than the other.[2]

While possibility is an objective feature of a situation, conceivability, as I construe it, is a relation between a situation and a cognitive subject. S is conceivable for X just in case there is at least one formally consistent representation of S

by which X represents S, and X is unaware of there being any formally inconsistent representation of S. When I judge that S is possible, the content of my judgment is a fact about S, that there does not exist a formally inconsistent description of S, not just that I don't know of one. But what makes it true of the pair, S and me, that S is conceivable for me, is a fact that involves me, that I don't know of any formally inconsistent description of S.

On this way of understanding the relation between possibility and conceivability, conceivability emerges as our only guide to possibility (for non-actual situations, that is), but not as a guarantee of possibility. What better evidence could I have that there is no formally inconsistent description of a situation than the fact that I don't know of one? On the other hand, there could be such a description nevertheless, in which case my judgment is mistaken. In this case, what is conceivable for me is not in fact possible.[3]

If modal intuition is indeed sensitivity to logical form, then it's important to determine which representational system's formal structure is at issue. For convenience I will speak as if it's the logical form of sentences in natural language that concern us, but in fact I think what determines our modal judgments is sensitivity to the logical forms of mental representations. Of course this commits me to the existence of a system of mental representation, as well as the further claim that it has a definite logical syntax. I'm happy with these commitments since I think they are necessary anyway to explain all sorts of facts about our mental lives.[4] I do not claim, however, that the precise nature of our mental logical syntax is transparent to us. Quite the contrary, it is undoubtedly a very difficult matter to discover the appropriate canonical notation for expressing it. Nevertheless, it seems to me that when I divide up situations into the possible and the impossible, what I'm doing involves, at least in part, a sensitivity to the logical forms with which I'm representing the situations in question.

Perhaps we can think of modal intuitions, as I'm understanding them, as akin to grammatical intuitions, as understood by the Chomskian linguist. The linguist posits a rich system of rules underlying our use of language. Intuitions, or judgments of grammaticality are reflections of this underlying competence, though an adequate description of the system of rules is itself a matter of empirical discovery. Similarly, modal intuitions, or judgments, reflect features of the logical syntax of thought, though it is a matter for empirical discovery to determine an adequate description of it.

Let's take some familiar examples to see how the framework applies. It's inconceivable that it should be both raining and not raining at the same time. This judgment reflects our capacity to detect the formal inconsistency in at least one way of representing the situation. It doesn't matter that there are ways of representing it that are not formally inconsistent. So long as we can apprehend at least one inconsistent description of the situation, we judge it to be impossible.

As another example, consider Kripke's (1980) arguments to show that proper names are not equivalent to definite descriptions. He relies heavily on our modal intuition to the effect that Aristotle could have failed to be a philosopher. If

"Aristotle" meant something like "the great ancient Greek philosopher who...", then there would be a description of the situation that was formally inconsistent, viz: "The great ancient Greek philosopher who...was not a philosopher". But, we don't judge this situation to be impossible, hence the description above does not count as a legitimate description of the situation. Therefore, "Aristotle" doesn't mean "the great ancient Greek philosopher who...".[5]

As a final example, consider a case where we originally thought a situation to be possible and then come to realize that it isn't. Before the advent of modern chemistry, it was thought possible that water is a simple, primitive substance, in fact one of the basic elements out of which everything else is constituted. But once we learned that it in fact has a molecular structure, we no longer think it possible that it be simple, or basic. What has happened? Originally, before the development of chemical theory, we only knew of representations of the situation, water's being simple, that were consistent. However, now we have discovered another description of that situation—"the substance with molecular structure H_2O is without internal structure"—that is inconsistent. So, it is no longer conceivable that water be a simple substance.

But the issue is a bit more complicated. So far, we've only been dealing with metaphysical possibility, and not epistemic possibility. Also, and I think this comes to the same thing, there is a use of "it is conceivable that..." which seems to apply to situations like the water example above even after we discover that water is H_2O. So even though we recognize (assuming we accept the Kripke-Putnam story about natural kind terms and proper names) that water is H_2O in every possible world, still we say that there is a difference between conceiving of water as a simple substance and conceiving of the weather as both raining and not raining. The former is epistemically possible, or still conceivable in some narrow sense, while the latter is neither. We need an account of this sort of conceivability, since it is plausible that it plays a major role in anti-materialist conceivability arguments.

I would accommodate this added notion of possibility this way. When originally considering a modal question—is this situation possible?—we always begin from an initial representation of the situation. To consider something is (at least in part) to represent it to ourselves, and this entails that there is some particular representation with which the consideration is accomplished. As per above, the question of possibility is this: this situation which I represent thusly, does it have any description (including this one, of course) which is formally inconsistent? If I can't think of one, then I judge the situation possible. But, after some thought, I might come upon the defeating description, an alternative description of the same situation which I apprehend to be inconsistent.

Now, there are two routes by which I might come upon this defeater: an a priori one and an a posteriori one. In the case of water, the route is clearly of the latter type. But sometimes, the route is an a priori one. Someone asks me if it's possible for there to be a married bachelor, and I say no. Of course the statement "There are married bachelors" is not itself inconsistent. However, I can determine a pri-

ori that it entails the statement "There are married unmarried persons", which is inconsistent. To reflect this distinction, I introduce the term "conceptual possibility". Situation S, described by R, is conceptually possible relative to R just in case there is no a priori entailment from R to a representation R' that is formally inconsistent. (For ease of exposition, I'll sometimes speak of statements themselves as conceptually possible or impossible.) Conceptual possibility is thus, as its name suggests, a feature of situations as conceptualized, or described, in a certain way.[6]

The difference then between the bachelor and water cases is this. Both situations involving married bachelors and situations involving water made out of XYZ are metaphysically impossible. In both cases there are descriptions of those situations that embody formal contradictions. However, we can capture the difference with respect to epistemic possibility this way. The statement, "There are married bachelors" is also conceptually impossible, but the statement "There is water made out of XYZ" is not.[7]

Before proceeding to discuss the mind-body case, let me address one possible concern with my account of conceptual possibility. Someone might object that by this definition I am committed to an analytic-synthetic distinction, as well as the existence of the a priori, both quite controversial notions. In fact, I am not so committed. If Quine is right and there is no a priori knowledge, and there are no analytic truths, then the category of the conceptually possible reduces to the class of consistent representations. This result follows from the definition of "conceptual possibility", since there will not be any a priori entailments from any consistent representation to an inconsistent one. This seems to me the right result for those who deny the existence of the a priori and the analytic. It turns out on their view that it is conceptually possible for there to be married bachelors. Fine, that's what it means to deny that anything (except logic)[8] is knowable a priori or that there are any analyticities. If you don't like that consequence, then you must believe in at least some a priori, analytic truths. The point is, my framework entails no stand on this question. It just allows a space for analyticity and a priori inference if such there be. However, I will have a lot more to say about this issue below.

2.

Now let's apply this framework to the anti-materialist conceivability argument. I will confine my attention here to qualia, leaving intentional properties out of consideration. I want to avoid many of the sticky questions about externalism that come up in discussions of intentionality.[9] But more important, it is with qualia that the conceivability argument poses its strongest challenge to materialism.

It is generally conceded that materialists are committed to a fairly strong metaphysical supervenience thesis. No two metaphysically possible worlds can agree on all their distributions of fundamental physical properties but differ in their distributions of mental properties. If we leave problems about certain relational

properties aside, then an even tighter principle is plausibly constitutive of materialism: no two creatures that are physically identical can differ mentally. Sometimes this is put in terms of the metaphysical impossibility of "zombies", creatures that are physically (or functionally, depending on one's version of materialism) identical to me but lacking in consciousness.

Assume S is the situation constituted by the existence of a zombie, and "P1...Pn" stand for our standard designations of the physical properties instantiated by a normal human being when experiencing qualia. Then the anti-materialist conceivability argument can be framed as follows:

(CP-Conceivability Premise) Relative to the description, "x has P1...Pn but not qualia", S is conceptually possible; that is, there is no a priori derivation of "x has qualia" from "x has P1...Pn".
(PP-Possibility Premise) If S is conceptually possible relative to the standard description, then it's metaphysically possible.
Therefore, S is metaphysically possible, which contradicts materialism.

There are two obvious ways for the materialist to respond. First, she can dispute CP, arguing that once we fill in actual predicates for P1...Pn, it will turn out that there is an a priori derivation from "x has P1...Pn" to "x has qualia". I mention this response only to set it aside. First of all, I don't find this bet on the future at all plausible, given what we already know of the sorts of physical and computational properties likely to figure in the group P1...Pn. Many materialists, though not all, would agree.[10] Secondly, since there is such a large group of materialists who do agree, I want to explore the implications of the other major response.

The second response is simple. Dispute PP. "P1...Pn" refer to the same properties that "qualia" and its more determinate versions (such as "reddish", "painful", and the like) do, but this can't be discovered a priori; it's a matter for empirical discovery. Given our characterization of metaphysical possibility, it turns out that zombies are not possible after all. There is a description of S that is formally inconsistent, namely "x has P1...Pn but doesn't have P1...Pn". This alternative description of S is not derivable a priori from the original description, so S is conceptually possible relative to the original description. But it is metaphysically impossible all the same.

It is traditional at this point to draw an analogy between the mind-body case and other examples of so-called "a posteriori necessities", for instance theoretical identities like "water = H_2O". While "the cup is full of water" doesn't follow a priori from "the cup is full of H_2O", still the conjunction of the first with the negation of the second (and vice versa) represents a metaphysically impossible situation; though conceptually possible relative to that specific description.[11]

If this were all there was to the anti-materialist conceivability argument, then it wouldn't be worth more than a moment's consideration. (Undoubtedly some in fact feel that way.) However, there is an important line of argument that constitutes a challenge to the standard materialist line on conceivability.[12] The idea is

this. Suppose two representations, A and B, pick out the same object/property, or necessarily related properties, but the identity or necessary relation is not determinable a priori. What explains this? It must be that there are two contingently related properties by which we conceive of the object/property(ies) in question, and these distinct conceptions are expressed by A and B respectively. Another way to put it is this: A's and B's "modes of presentation" involve distinct, contingently related properties.

So, for instance, take the identity of water with H_2O. The reason it takes an empirical discovery to learn this identity is that this substance instantiates many contingently related properties and it takes empirical work to discover that it's one and the same thing that instantiates them. On the one hand it possesses a certain molecular structure, and on the other it manifests certain superficial features, such as liquidity at room temperature. Similarly, the planet Venus instantiates both the property of appearing at a certain location in the sky in the morning and also the property of appearing at a certain location in the sky in the evening. It is the distinctness of the properties through which we represent the objects/states in question, and the fact that these properties are only contingently related, that explains our inability to determine a priori that the objects/states are one and the same.

Let's call this model of explaining a priori ignorance of an identity (or necessary relation) the "distinct property model" (DPM). One might wonder whether the model fulfills a genuine need; that is, whether the phenomenon of a priori ignorance of a necessary relation is puzzling to begin with. I think we can motivate the puzzle this way. Suppose it true that properties P and Q are necessarily related, so there is no possible world where you can have P instantiated without Q (and/or vice versa), yet the statement "x has P but not Q" is conceptually possible, and therefore its negation is knowable only a posteriori. Well, one way to characterize empirical information is by reference to possible worlds: empirical information locates our world within a particular region of the space of all possible worlds, thus distinguishing it from those possible worlds lying outside that region. So now one might well ask how it could be necessary to secure information that distinguishes our world from other possible worlds in order to determine a fact that holds at every possible world.

There seem to be two obvious answers: the DPM and one that doesn't work.[13] The one that doesn't quite work is this. The reason you can have a conceptually possible statement that doesn't represent a metaphysical possibility is that logical form is sensitive to vocabulary, and so long as two different non-logical vocabulary items can represent the same object or property, such situations will arise. An obvious example is the water-H_2O case. "Water is water" is conceptually necessary, while "water is H_2O" is not, and the difference is clearly that "water" and "H_2O" are different terms.

But this answer is inadequate. We don't normally refuse the honorific "a priori" to a statement merely on the basis that someone might not know a crucial vocabulary item. That bachelors are unmarried doesn't count as knowable a pos-

teriori merely because I have to learn what "bachelor" means before I would assent to it. So, a constraint on an adequate explanation of the a posteriori status of many identity statements is that it not appeal to the subject's incompetence with any of the crucial terms or concepts.

But now we really do face a puzzle. How could we know what both terms mean, or, less tendentiously, be competent with both terms, and yet fail to know that their referents are necessarily connected (or co-refer)? How could it take empirical experience to fill the gap in our knowledge? Well, if there are possible worlds where their referents are not connected (or they don't co-refer), even while meaning what they do, then of course it would take empirical experience to determine that this wasn't one of those worlds. But how could this be?

This is where the DPM comes in. We satisfy the above-mentioned constraint by identifying meaning with mode of presentation, and then add that for many terms reference is determined by a combination of mode of presentation and contextual features of one's use and/or acquisition of the term. Since the same mode of presentation can pick out different objects or properties in different contexts, we can see where the need for empirical knowledge comes in. We need to know which among the possible worlds, considered as contexts, our world is, for in some contexts the referents of the two terms will not be necessarily connected, or identical. This is why it is crucial that the properties involved in the modes of presentation of the two terms be only contingently related. For if they were necessarily related, it wouldn't matter which world-context we were in, since they would pick out necessarily connected or identical referents in every world-context. Thus the DPM seems to be the only way to go.

If we apply the DPM to the case of pain, then it looks as if the materialist is in trouble again. The materialist's strategy was to deny PP, citing the possibility of empirically discovering identities of the form "Pain = the firing of C-fibers". But suppose, for the sake of argument, we grant this. Still, in order to explain why this identity can't be discovered a priori, it is necessary to posit distinct, contingently related properties, say "painfulness" (a qualitative property) and the property of having one's C-fibers firing. But now we can just run the entire argument on painfulness itself which, by hypothesis, is only contingently related to the firing of C-fibers. The problem is that whatever physical description we substitute for "C-fiber firing" as a candidate for identity with painfulness, we will always face the following dilemma: either we have to deny that the identity in question is empirical, or we have to admit a new, contingently related mental property to serve as the mode of presentation for the mental term. Since we're committed to the empirical character of these identities, we're stuck. We seem committed to what Smart's (1959) objector called "irreducibly psychic" properties after all.

As I'm interpreting it currently, the anti-materialist conceivability argument depends upon acceptance of the DPM. So far we've seen one argument for the DPM; that it is needed to account for how we could be ignorant of certain metaphysical necessities while yet knowing the relevant meanings. In addition to this

mixed epistemic-semantic argument, I'll consider two other epistemic arguments in section 3. But for the remainder of this section, I'll consider an alternative to the DPM, one that allows the materialist to block PP without introducing irreducible mental properties yet again.

One might have already been made suspicious by the DPM's heavy reliance on the existence of a priori, analytic connections between concepts.[14] Those moved by Quine's attack on the a priori and analyticity won't look kindly on the DPM. However, the DPM's defenders are often undaunted by Quine's attacks, and I think their explicit positive arguments need to be addressed. Fortunately for the materialist, I think they can be.

To begin, let's distinguish two kinds of mode of presentation (where by a "mode of presentation" is meant the means by which a representation connects to its referent): ascriptive and non-ascriptive. An ascriptive mode is one that involves the ascription of properties to the referent, and it's (at least partly) by virtue of its instantiation of these properties that the object (or property) is the referent. A non-ascriptive mode is one that reaches its target, establishes a referential relation, by some other method. The object isn't referred to by virtue of its satisfaction of any conditions explicitly represented in the mode of presentation, but rather by its standing in some particular relation to the representation. The mode of presentation is the relation itself. The usual candidate for such a relation is some causal or nomic relation, such as covariation between representation and referent that meets certain constraints.[15]

There are good reasons for thinking that if you want to avoid meaning holism, then at least some modes of presentation will have to be of the non-ascriptive sort. But once we admit that not all modes are ascriptive, it becomes an open question whether even the standard examples really do exemplify ascriptive modes. So again consider "water". On some atomistic views of content, "water"'s reference is determined by causal factors that, as it were, work "behind the scenes" to establish the referential relation. It may be true that one's central beliefs about water, that it is liquid at room temperature for instance, might contribute to sustaining the requisite causal or nomic relation between the term and its referent, so that were one to abandon enough of these beliefs the requisite nomic connection would be broken. Still, it isn't the case that these beliefs determine reference by constituting satisfaction conditions for the term in the relevant context. The mode of presentation in such a case is constituted by the term's causal connection to water.

As I said, some philosophers argue that the best view of the standard natural kind cases is that their modes of presentation are non-ascriptive.[16] The basic reason is that they feel this avoids problems about drawing an analytic-synthetic distinction or falling into holism. Though I'm quite sympathetic to this position, it's not necessary for my argument here to positively endorse it. Rather, the point I want to make is that the availability of such modes undermines the anti-materialist conceivability argument by obviating the need for the DPM to explain a priori ignorance of necessary connections.

Suppose we have two terms, A and B, at least one of whose modes of presentation is non-ascriptive; say it's A's. Let's assume that A and B both refer to O, but we can't know this a priori. The question we posed above was this: if A and B refer to the same thing, what space could there be for empirical information to play a role in determining this fact, aside from the need to learn the terms themselves? On the DPM, the answer was that A and B both expressed modes that ascribed properties that were only contingently related to each other. With non-ascriptive modes in play, we can now propose an alternative to the DPM, call it the Revised Distinct Property Model (RDPM). We still appeal to distinct properties associated with the two modes of presentation, but we drop two crucial conditions: (1) they don't have to be contingently related, and that's because (2) they don't have to be explicitly represented within the mode. In fact the subject need not know of the relevant property at all.

The idea is this. Suppose, to over-simplify, A's mode of presentation is constituted by something's being the cause of A-tokenings. We can suppose also, along with the materialist, that a full enough description of the relevant physical facts will metaphysically necessitate O's instantiating the property of being the cause of A-tokenings. The puzzle was how this bit of necessary truth could be unavailable to the subject competent with both the relevant physical descriptions and A. The answer is that the subject can be competent with A without knowing the crucial fact that to be A's referent an object must be the cause of A-tokenings. But how could a subject competent with A, one who knows A's meaning, be ignorant of the very property that constitutes the mode of presentation it expresses? That's how non-ascriptive modes work. They establish relations, again, "behind the scenes", not by being cognitively grasped by the subject. The subject's competence with the term, her "knowledge" of the meaning, consists entirely in her instantiating the requisite relation to something in the world. A priori ignorance is thus attributable to ignorance of what constitutes the meaning in this case.[17]

Once we allow the possibility of non-ascriptive modes of presentation for qualia, the inference from the conceptual possibility of zombies to their metaphysical possibility is undermined. The standard materialist strategy of accepting CP but denying PP can work. There is indeed an inconsistent description of the zombie situation, but it can be derived from the initial description only with the addition of identity statements that are not knowable a priori. This is no problem since the a posteriori status of these psycho-physical identities does not entail the existence of distinct, contingently related properties to serve as the modes of presentation for the terms flanking the identity sign. Thus, there need not be metaphysically irreducible mental properties after all.

3.

In this section I want to consider two further epistemic arguments for the DPM. I'll call these the "argument from knowledge of identity" and the "argument from

explanation". In a sense they are both transcendental arguments. Though it is not incumbent on the materialist to deny that natural kind terms like "water" express ascriptive modes of presentation, it is crucial that the arguments we are considering not force her to accept this either, since the very same arguments would apply to qualitative concepts. So we'll continue to use natural kind terms as our testing ground.

The argument from knowledge of identity goes like this.[18] Without appeal to the analytic connections involved in the DPM we can't make sense of the standard cases of allegedly synthetic, a posteriori judgments of theoretical identity in the first place. So, take the case of water. What justifies us in claiming that water is indeed identical to H_2O, if not a prior—i.e. an a priori—understanding that by "water" we mean "the local watery stuff", together with the empirical discovery that H_2O fits the bill?

In reply, it doesn't seem to me that a priori, analytic connections are necessary to justify our theoretical identity judgments. Suppose "water" has a purely non-ascriptive mode of presentation. It might seem that we lack an epistemic handle with which to connect it now to H_2O. But notice, first of all, that even if this were true it wouldn't affect the actual identity of water with H_2O, just our ability to know the identity. Secondly, to get the required "epistemic handle" it isn't necessary for us to have a priori access to anything about either water or "water"; it's enough that we have fairly well-justified beliefs about water that are expressed with "water". That I believe very strongly that water is the local watery stuff is not in question here. Since I do, and since I discover that H_2O is also the local watery stuff, I conclude that they are the same thing. No a priori knowledge is necessary.

The DPM advocate might object that while it's part of the justification of the identification of water with H_2O that one's beliefs about water turn out to be about H_2O, there is still a choice left to make that the empirical discovery doesn't touch: namely, whether water is to be identified with the water-role itself or with its occupant. To make this decision we have recourse to our modal intuitions, which involve judgments about what we would say under certain circumstances, such as the Twin Earth example, and this reflects a priori knowledge concerning our concept.

Undoubtedly the objector is right that there are two issues here. First, there is the question, answerable by empirical means, about which property/substance plays the water-role. As far as the metaphysics are concerned, both properties exist: the property of filling the water-role, and the property of being H_2O. Second, there is a semantic question: given the existence of both the role property and the occupant property, to which one do we refer with our use of "water"? The fact that, on reflection, we would not call XYZ "water", does seem to show that we are using "water" to refer to the occupant, not the role. Does this show that "water" has an a priori analysis?

No, not really. Again, as I argued with modal intuitions earlier, we can treat semantic intuitions—what we would say in various circumstances—the way the syntactician treats grammaticality intuitions. By considering what we would say

we garner evidence for the correct semantic theory. The fact that we are inclined not to call XYZ "water" reveals to us, as evidence bearing on an empirical hypothesis, that our concept of water is of a role-occupant, not a role itself. But our ability to reflect on our practice is not itself constitutive of the practice. Therefore, we still have no argument that a priori knowledge of a certain sort is necessary for concept possession.

One way to see this is to note the possibility of concept possessors who lack the reflective capability we have. Higher animals present one sort of example. Lacking such a capacity—to entertain hypothetical cases and render judgments about what we would say—would certainly be a hindrance for constructing a semantic theory, just as it would be for constructing a syntactic theory. Nevertheless, it wouldn't show that there was nothing for such a theory to be about.[19]

The defender of the DPM might try the opposite tack; rather than arguing from the possibility of empirically discovering certain identities, arguing instead from the conceptual impossibility of discovering certain other identities. For example, can one conceive of water turning out to have none of the standard properties by which we recognize it? Could scientists tell us that this stuff we drink, falls from the sky, fills lakes and oceans, isn't really water after all? Not, mind you, that some of it isn't.[20] Rather, what I'm talking about is a case where none of the stuff we pretheoretically picked out as water turns out to be H_2O, but the scientists tell us that water is H_2O nevertheless. Certainly in such a case we would say they were wrong. By "water" we meant this stuff in the lakes and oceans and the stuff we drink. If you're not talking about *that* when you identify it with H_2O, then you're not talking about water. But what could explain the strength of this conviction if not possession of a priori, analytic knowledge of the meaning of "water"?

In response, one can maintain the denial of analytic connections even in the face of this argument, so long as one has an account of how certain beliefs about water can be privileged in some other way. I don't think it's hard to see how to do this. Clearly there is some surprising stuff I'm prepared to find out about water as a result of scientific investigation. It's enough to note in this connection an example Ned Block[21] is fond of, that glass turns out to be a liquid. But surely not all of our central beliefs, especially those that are most closely tied to the circumstances in which we apply the term/concept, could be wrong; for what would then tie our concept to the property it's alleged to be about? This isn't an appeal to analyticity, but rather an appeal to the empirical conditions necessary for maintaining whatever causal or nomic connection constitutes reference.

Furthermore, remember that what normally justifies an identification of a common property/substance, such as water, with a scientifically discovered property, is the use to which that identity statement can be put in constructing explanatory arguments whose conclusions express the commonly-held beliefs about it. So, an identification which contradicted every single one of these commonly-held beliefs would be hard-pressed to find any justification. Again, this consideration makes no appeal to analytic connections. Rather, the inconceivability of the situation in which all of our commonly-held beliefs about water are false is ex-

plained by the strength of our conviction that no theoretical identification entailing such a consequence would be justified.

This brings us to the second epistemic argument, the argument from explanation. I've argued that it is not a requirement of our possession of the concept water, or understanding the term "water" that we also possess some conceptually connected description of its superficial properties, a description of the water-role. Yet, we do in fact possess such a description, whether you want to call it conceptually connected or not, and it clearly plays a crucial role in the epistemology of chemical discovery. In particular, it's hard to see how we could explain anything about water by appeal to its chemical composition, unless two conditions were met: first, there were descriptions of various superficial properties of water that stood in need of explanation, and second, we could derive these descriptions from the descriptions of water's chemical composition together with various chemical principles and laws.[22]

For instance, I want to know why water is liquid at room temperature. The story goes roughly like this. Room temperature is a state of matter constituted by a certain mean molecular kinetic energy, call it 'r'. H_2O molecules, when at r, are bonded in such a way that they display the motion syndrome constitutive of liquidity. For this to really constitute a full explanation, a description of the liquid state of water at room temperature should be formally derivable from the descriptions of the properties of H_2O molecules at r. If we couldn't, at least in principle, turn this explanation into a genuine derivation, then for all we would know it would be in fact possible to have H_2O molecules at r (and have the rest of the relevant chemical facts stay the same) without water's being liquid at room temperature. But if this is possible, then we still don't know why water is in fact liquid at room temperature. It may well be that the macro facts supervene on the micro facts, even though we can't derive a description of the macro facts from the micro facts. But this would be small comfort, for we would have lost the explanatory power we expected from the execution of the materialist program. Thus, to the extent we feel that we do indeed have a genuine explanation of the liquidity of water at room temperature, a derivation of the macro facts from the micro facts must exist, and be (in principle) accessible to us.

Until now we've seen no argument that committed the materialist to an a priori connection between what's expressible in physical vocabulary and what's expressible in non-physical vocabulary. But the existence of successful explanations seems to provide just the connection the DPM advocate needs. For now she can ask how such derivations are possible if it isn't the case that one's grasp of the concepts in the explanandum consists at least in part in knowledge of analytic connections to other concepts.

But in fact this argument fails as well. There is indeed a problem concerning explanatory derivations that appeal to analytic connections can solve. The problem is this. Given the difference between the vocabularies in which the micro facts and macro facts are expressed, how do we get a derivation of the latter from the former? Well, the defender of DPM says this. All of the terms used to describe

the macro facts—from "water" to "liquid"—have analytic connections to descriptions of causal roles which themselves involve only quantifiers and terms that are held in common with the micro descriptions (such as mathematical quantities, specifications of spatio-temporal locations, and the like). These descriptions of causal roles, given their analytic equivalence to the originals, can then be substituted for all the macro terms appearing in both the explanans and the explanandum. Once the substitution is completed, the problem of disparate vocabularies, which seemed to present an obstacle to a derivation of the explanandum from the explanans, vanishes.

The account just presented certainly provides one way of understanding how the requisite derivations are possible. However, there is an alternative account available. Suppose we just took the relevant identities as empirical premises in the derivation, your standard bridge principles. So, we have that water = H_2O, that liquidity = a certain property specifiable in micro and spatial terms, etc.. In fact, this very straightforward, simple solution to the problem fits well with our answer to a question that arose earlier; namely, what justifies the claim that water is H_2O in the first place? According to the advocate of the DPM, this claim is the conclusion of a derivation that contains the analytic definition of "water" in terms of the occupant of the water-role together with the empirical premise that H_2O in fact occupies the water-role. But the opponent of the DPM presents quite a different picture. That water is H_2O is not the conclusion of any derivation. Rather, it functions as a premise in various explanatory arguments which have descriptions of water's macro properties as their conclusions. When asked for the justification of the premise itself, the answer is that it's justified because of the explanatory role it plays. By accepting the claim that water is H_2O it's then possible to show why water has the superficial properties it has. No analytic definition need enter the reasoning either to support the identity claim itself or to function in the various explanatory arguments from the micro-chemical facts to the macro facts about water.[23]

Notice that the micro-explanation of the macro-fact that water is liquid at room temperature contains three bridge principles among its premises: that water = H_2O, that liquidity = a certain spatial behavioral syndrome, and that room temperature = mean kinetic energy r. If asked for an explanation of any of these three identities the correct response is to express perplexity about what it means to explain an identity anyway. Things are what they are; there is no sense to explaining that. What we might be asking for is evidence of the identity, a question that can look awfully similar to an "explanation-seeking why question", as in "why should water be H_2O?". But the evidential question is answered by pointing to explanations of other facts, such as the fact that water is liquid at room temperature, which depend crucially on acceptance of the identity of water and H_2O.

If the materialist reply to the argument from explanation holds up, then it really undermines the anti-materialist position quite severely. For what seems most plausible about the anti-materialist's case is that the conceivability of zombies reveals an "explanatory gap" when it comes to consciousness that is not

present when dealing with other macro phenomena.[24] Yet, if we apply the considerations just presented to the case of qualia, it's not clear how to make this case.

Take my current visual sensation as I focus my attention on my red diskette box. Call that token sensation r. I'm also in a certain brain state that corresponds in some way with r, call it b. Now, many philosophers, even anti-materialists, would agree with identifying r and b. These are the same token state. So we have (1) r = b.

Okay, let's consider the state types of which r/b is a token. We can think of these as properties, being a state of type R and a state of type B, where "R" stands for the reddish qualitative character of the visual sensation, and "B" stands for its neurophysiological character.[25] Many philosophers, again even anti-materialists, would agree with the claim that every token of R is a token of B, and vice versa, at least as a matter of law. So we have (2) It is at least nomologically necessary that $(\forall x)(Rx \equiv Bx)$.

The real question, then, is whether to go ahead and identify properties R and B. Should we adopt (3) R = B (taking "R" and "B" here as singular terms referring to properties)? Let's see how this question comes up in the context of seeking an explanation. One explanatory question I could ask is this: why am I in state r when I look at the diskette case? Answer: there's a physical story which starts from the light reflecting off the diskette case and ending with my occupying state b. If we adopt the hypothesis that r = b, then I can explain why I'm in r, so I have good reason to accept (1), and I have my explanation.

Of course the anti-materialist will insist that this isn't the relevant question. The question she wants an answer to is this: why does the state I'm in have property R? Here's one answer: Since the physical story explains why it has B, and since we already accept (2), it follows that it will have R. The anti-materialist will of course not accept this dodge, but will press further: okay, what explains (2)?

If (2) doesn't have an explanation, then that can be for only one of two reasons: either we've reached a basic law, or it isn't a matter of law at all, but just follows from (3). With respect to the first alternative, the point is this. We recognize that some laws of nature are just brute, fundamental facts about our world, and others are determined by these basic laws. The anti-materialist is willing to admit the basic nature of (2) as a reason for its lacking an explanation, but then insists that admitting (2) to the class of basic laws is really to abandon materialism. I think many materialists would agree with that, and so wouldn't want to take this way out.

So now we see the importance of (3). Can we use (3) as we used the identities of water with H_2O and liquidity with its peculiar spatial behavior syndrome? Suppose we argued like this. By adopting (3), I can explain (2). The reason every instance of B is an instance of R is because R and B are the same thing. That's what explains the correlation. Furthermore, you can't go on to ask why (3) is true, because, as we claimed above, identity facts, unlike law facts, aren't the sorts of

facts that stand in need of explanation. Things are what they are. You can ask why one should think that the identity statement is true, of course, and this is just to ask for one's evidence. In this case, (3)'s ability to explain (2) would be the answer. But you can't ask what, metaphysically speaking, makes it true. That question has no sensible answer.

Let's take stock. The anti-materialist conceivability argument went like this. Materialism is the thesis that the fundamental physical facts metaphysically determine all the facts. By MBP, if one set of facts metaphysically determines another, then there must exist a description of both sets under which the determining set formally entails the determined set. By CP, with which we are assuming materialists agree, zombies are conceptually possible relative to the standard representation of the fundamental physical facts and the facts about qualia. This means that there is no formal derivation of the standard description of the facts about qualia, nor any other description derivable a priori from the standard description, from the standard description of the fundamental physical facts. So, any derivation of the standard description of the qualitative facts from the standard description of the fundamental physical facts must include a posteriori identity statements among its premises. By the DPM, all a posteriori identity statements involve the representation of distinct, contingently related properties through the modes of presentation expressed by the terms flanking the identity sign. Thus there must be some qualitative properties that are only contingently related to physical properties. Hence, not all the mental facts are metaphysically determined by the fundamental physical facts.

The materialist replied by attacking the DPM. If the DPM is false, and there are non-ascriptive modes of presentation, then no new properties need be introduced to explain the a posteriori nature of the requisite bridge identities. As support for her rejection of the DPM, the materialist showed how we could account for our competence with standard kind terms like "water", justify the a posteriori identification of water with H_2O, and explain whatever needs explaining about water, all without recourse to analytic-a priori connections between "water" and any topic-neutral description of the water-role. The upshot is this. Just as zombies are conceptually possible relative to the standard descriptions, so is "zombie-H_2O" (i.e. H_2O that isn't water, even given all the relevant micro-physical facts). If the conceivability hypothesis amounts only to the conceptual possibility of zombies, then it poses no threat to materialism.

4.

Confirmed materialists can stop here, secure in the conviction that their doctrine is safe from the conceivability argument. Yet many philosophers, myself included, will feel that something is still wrong. In particular, they may still feel that there is an important distinction to be made with respect to explanatory adequacy between the water case and the qualia case. This distinction in explanatory adequacy seems connected with a distinction between the senses in which

zombies and zombie-H_2O are conceivable. In this section I want to do two things. First, I will present an argument to support the feeling that there is an important distinction between the case of qualia and other cases of empirically grounded identities. Second, I'll present a new version of the anti-materialist conceivability argument, one that does not rely on the DPM, and therefore is not vulnerable to the materialist counter-attack presented in sections 2 and 3. In section 5, I'll consider how the materialist might reply to this new argument.

The materialist's response to the argument from explanation in section 3 relied crucially on the premise that an identity is not the sort of fact that stands in need of explanation. Of course there are identity claims that one can seek explanations for, but they always turn out to really be, if not requests for evidence, questions about how or why distinct properties are coinstantiated. So, for instance, I can express wonder that this full-grown man I am now facing is the same person as the little boy I met 20 years ago, or even that this apparently continuously divisible liquid I call "water" could be the same thing as a collection of H_2O molecules. But in both these cases it's clear that what I'm wondering about is how the very same object could instantiate these very different properties. To wonder about pure identities, how X could be itself, where no distinct properties are involved, doesn't seem intelligible.

Yet, when we look more closely, it seems that things are not quite so straightforward. In particular, there is a sharp epistemic contrast between various standard cases of identity claims and the case of an identity claim like (3) R = B above. With the standard cases, once all the relevant empirical information is supplied, any request for explanation of the identities is quite unintelligible, as the considerations just adduced would predict. In the case of (3), however, it seems quite intelligible to wonder how it could be true, or what explains it, even after the relevant physical and functional facts are filled in. This difference calls out for explanation.

To illustrate the distinction I have in mind, consider three cases: one involving natural kinds, one involving indexicals, and one involving demonstratives. Take indexicals first. It's notorious that no purely descriptive, or qualitative statement can entail one containing an indexical.[26] No matter how full a description there is of what's happening to a certain body in a certain spatio-temporal location, it won't logically follow that it's happening to me here now. Yet, of course I can explain why I have a cut on my hand now by citing events describable in terms like "this body at time t encountered a knife while cutting a bagel...". I don't get a derivation, of course, without bridge principles, like "that body is mine" and "now is soon after time t", etc.. But neither the materialist nor the anti-materialist will claim that the inability to derive "My hand is cut now" from non-indexical premises shows the metaphysical irreducibility of the properties, being mine and being now.

More significant than the issue of metaphysical irreducibility, however, is the question of the epistemic status of the bridge principles. After being informed that the body referred to is mine, and the time t is shortly before now, would there

be any sense to the question, "how could that be my body?" or "how could t be now?"? Of course I can think of circumstances that might give sense to these questions, but they involve ascribing different properties to, say, the body referred to in the premise and those I believe my body to have. What there really doesn't seem to be is any cognitively significant content to the notion of being mine, pure and simple. "Mineness", as a property in its own right, is not a genuine property in the same sense as height, weight, and appearance. Thus to wonder how the body referred to in the explanans could have it (i.e. "mineness") doesn't seem to have any cognitive substance.

Similarly, take the following example, involving a demonstrative. I point blindly in front of me and say, "I wonder what that is?", having no more substantive idea of what I'm pointing at than that it's an object occupying space. I now open my eyes and see that the object in the line of sight from my pointing finger is my red diskette case. Is there any sense to be made of my now wondering: but how could my red diskette case be *that*? I don't see any. Once I've determined that it's the red diskette case that occupies the relevant contextual niche, there's just nothing more to wonder about.

Notice that there is still no way of deriving the statement, "That = Joe's red diskette case" from premises containing only non-demonstrative terms. Somewhere along the derivational route one needs to encounter a premise to the effect that *that* is what Joe is pointing at. So the lack of a demonstrative-free derivation is not sufficient to show that there is still some sense to wondering how *that* could be my red diskette case.

Finally, turning to natural kinds, consider again the case of water and H_2O. Earlier we determined that there was no way to derive statements containing the term "water" solely from premises that did not contain the term. Yet, once we discover that H_2O is indeed the substance that lies at the other end of the contextual reference-determining relation from "water", it does seem that there is little sense to be made of my wondering how H_2O could be *water*. But what do you have in mind, one is tempted to ask of me? Of course I may answer that I don't see how H_2O could play the water-role, how it could be liquid, transparent, quench thirst, etc.. These questions do have sense, but then they also have an answer in terms of underlying chemistry. It's after all is said and done, the chemical explanations are all in place, and I still persist in my wonderment, that one is absolutely puzzled as to what substantive content there could be to my wondering. At that point it just seems as if I'm holding on to the word with nothing in mind that it signifies.

In stark contrast to these three examples stands the case of qualia. I am told that my concept of reddishness is really about a neurophysiological or functional property. I then wonder, as I ostend the reddishness of my visual experience, how could a functional or physiological state be *that*? In this case, even if one is convinced by the identity claim, one wouldn't be mystified as to what it is I'm wondering about. There does seem to be a substantive content to my puzzlement. Finding out that a particular neurophysiological or functional property stands at

the other end of the contextual reference-determining relation from my representation "reddishness" doesn't settle all there is to be settled, as it seems to with "water".

So, we are faced with the following contrast. Once all the standard superficial properties of water are explained by reference to the structure of H_2O molecules and general chemical laws, there seems to be no substantive cognitive significance to the question how water could be H_2O. On the other hand, even after all the causal role properties of experience are explained by reference to its neurophysiological or functional structure, still there seems to be genuine, substantive cognitive significance to the question how reddishness could be a neurophysiological or functional property.

It might be thought that the contrast between water and reddishness could be explained this way. In the case of water, we start out with a host of property-ascriptions along with the contextual feature. Again, it doesn't matter whether one incorporates the property-ascriptions into the meaning of "water", or just allows that whatever the cognitive content expressed by "water" itself, it's still the case that we possess a rich web of associated beliefs concerning it before scientific investigation gets off the ground. Thus, when scientific investigation yields a candidate to which these beliefs apply, we find the identification irresistible, and this explains the apparent lack of sense we find in questioning the identification. However, with "reddish" there isn't the same web of associated beliefs. Our primary cognitive contact with this property, when presented introspectively, is purely contextual. We just ostend it as we instantiate it. Thus there is bound to be a residual sense of inappropriateness about the suggestion that it is identical to some richly described theoretically posited property.

In reply I would note the contrast with our other examples, especially the "blind" use of "that" mentioned above. It seems to me that what the materialist is suggesting in order to explain the contrast between water and reddishness is that qualitative concepts are essentially "blind" demonstratives. They are pointers we aim at our internal states with very little substantive conception of what sort of thing we're pointing at—demonstrative arrows shot blindly that refer to whatever they hit. But just as it seemed unintelligible to wonder how water could be H_2O after learning the relevant chemistry, it is similarly unintelligible to wonder how my red diskette case could be *that* when I point blindly and am told that I've pointed at my red diskette case. If the materialist were right about my concept of reddishness, it should behave just like this case of blind pointing. But, as we've seen, it doesn't.

I'll call an identity claim that admits of an intelligible request for explanation a "gappy identity". Now, let's apply this distinction between gappy and non-gappy identities to the question of conceivability. In general, there is a tight connection between explanation and conceivability. An explanation is called for whenever it is conceivable that it could have gone the other way. I want to know why most objects fall when unsupported. My question is intelligible precisely because we can conceive of their not falling. But now let's explicitly draw a

distinction between two grades of conceivability in terms of the distinction between gappy and non-gappy identities. I'll call a situation "thinly conceivable" relative to R just in case it's conceptually possible relative to R. This is the sense of conceivability that was employed in CP of the initial argument presented in section 2. I'll call a situation "thickly conceivable" relative to R just in case it's conceptually possible relative to R, and any derivation we can construct from R to a formally inconsistent representation R', will include gappy identities in its premises.

At the end of section 3 I put the materialist reply to the conceivability argument this way: "Just as zombies are conceptually possible relative to the standard descriptions, so is "zombie-H_2O (i.e. H_2O that isn't water, even given all the relevant micro-physical facts). If the conceivability hypothesis amounts only to the conceptual possibility of zombies, then it poses no threat to materialism." However, now the anti-materialist can argue that there is a crucial difference. Granted, zombie-H_2O is conceptually possible, or thinly conceivable, and this poses no threat to the identification of water with H_2O. But zombies are not only conceptually possible, they're thickly conceivable, since the only way to derive a contradictory representation of a zombie is to use gappy psycho-physical identities as bridge principles.

In a way, the anti-materialist is using the materialist's own argument against her. The materialist argues that just as zombies (under the appropriate description) are conceptually possible, so too is H_2O without water, or the corresponding nonindexically described situation without its indexically described counterpart. Since we're not tempted to posit any irreducible properties there, or find any explanatory gaps, there's no reason to in the case of qualia. But the anti-materialist insists that it's of the utmost significance that we aren't tempted to find irreducible properties or explanatory gaps in the water or indexical cases, but we are in the qualia case. The difference requires an explanation. The best explanation available is that in the case of qualia we're dealing with genuinely independent properties.

We can now reconstruct the anti-materialist conceivability argument as follows:

(CP′) Zombies are thickly conceivable.
(PP′) Thickly conceivable situations are metaphysically possible.
Therefore, Zombies are metaphysically possible.

Parallel to the materialist response to the first version of the argument, the materialist might try to attack PP′ on the grounds that empirically grounded identities can show us that what is conceivable is not always possible. But this response is weaker in two ways than the attack on PP in the original argument. For one thing, the compelling analogy with the other empirically grounded identities breaks down over gappiness and thick conceivability. While we have nice examples of thinly conceivable situations that are nevertheless metaphysically impossible, we don't have any examples of thickly conceivable situations that are

metaphysically impossible. At least, so the anti-materialist argues, and it's hard to think of an example to prove her wrong.

Secondly, now that we're talking about thick conceivability, the anti-materialist can press a new argument for the applicability of the DPM (or some principle quite like it). Though we showed that the DPM was not needed to account for the a posteriori character of the standard theoretical identities, it is needed to account for the gappiness of psycho-physical identities. The argument goes like this. What explains gappiness? There seem to be two possibilities: either the identity is true, and what we need explained is how this one thing could instantiate some particular pair of distinct properties we believe it to have; or the identity is false. The point is that an intelligible request for explanation seems to entail a distinction in properties somewhere. If it isn't to be located in the properties of the one thing we're representing on both sides of the identity sign, then it must be that the terms flanking the identity sign themselves represent distinct things. Either way, we're left with irreducible mental properties.

Before turning to evaluate this new argument in section 5, I want to present one final argument in favor of there being an important distinction between the standard cases of empirically grounded identities and the case of qualia. This argument bears directly on the question of the alleged metaphysical independence of qualitative properties. Consider the problem of attributing qualia to other creatures, those who share only our functional organization, say, or maybe only our higher-level but not very fine-grained functional organization. I take it that there is a very real puzzle whether such creatures have qualia like ours, or even any at all. How much of our physico-functional architecture must be shared before we have similarity or identity of experience?[27]

The contrast here with the case of water is instructive. We are faced with XYZ and with an alien creature, and we have to decide whether XYZ is water and whether the alien creature experiences reddishness. In both cases, let us suppose, we have all the relevant information concerning physical structure and causal role. We know that XYZ is a different molecular structure from H_2O, and that on Twin Earth it plays the "water-role". With respect to the alien, we can suppose that we have a relatively complete map of its functional organization and the way that organization is realized physically.

Now, the questions are: is XYZ water? does the alien experience reddishness? Earlier I conceded to the defender of the DPM that the former question is essentially a semantic one. We know what there is to know about XYZ, we just have to determine whether the term "water", as used by us, applies to it. To decide this, I said, we consult our linguistic intuitions. If those intuitions don't determine an answer, then it seems quite open to say that it's now a matter for decision whether we should extend our application of the term "water" to XYZ or not. As it happens, I think the argument that our linguistic intuitions settle the matter are persuasive, but nothing of substance here hangs on that.

But when it comes to the second question—does the alien experience reddishness?—it clearly isn't merely a semantic question at all. I'm not asking whether

or not to extend my use of the term "reddishness" (or its mentalese equivalent) to the alien, and the answer to the question doesn't seem to lie in consulting my linguistic intuitions. I want to know whether or not it has this sort of experience, whether or not it instantiates a certain property. Furthermore, the idea that, failing to find sufficient grounds for answering the question either way, we might resort to just deciding whether to say it has reddish experiences or not seems preposterous. You don't just decide matters of fact. If one feels there really is a contrast here, as I do, then it seems to commit one to the claim that reddishness is a genuinely independent property. The point is that only if reddishness is distinct from either a physical or a functional property could there be more than a semantic question left to decide.

5.

As with the initial conceivability argument, there are various options open to the materialist by way of reply. First, she can reject the conceivability premise itself, (CP′). Now that we're talking about thick, as opposed to thin, conceivability, this move is much more plausible. Many philosophers who buy the thin conceivability premise, (CP), do so because they aren't comfortable with any claims involving a priori analyses of concepts. To the extent that thick conceivability goes beyond the mere denial of such analyses, they may very well fail to accept it. If one isn't moved by the contrast between the cases of water, demonstratives, and indexicals on the one hand and qualia on the other, this is the place to block the argument. But I take the argument of section 4 to reveal the existence of a genuine explanatory gap, a genuine distinction between gappy and non-gappy identities. Therefore, I'm interested in what metaphysical conclusions can be drawn if we accept CP′.

So suppose we accept CP′. Must we accept PP′? I want to explore two ways of resisting the inference to PP′. The anti-materialist is relying on two features of the DPM-style explanation of gappiness: that wherever there is gappiness there must be two distinct properties in play, and that the two properties in question must be only contingently related. One way to attack the argument, then, is to accept that there must be distinct properties in play, but allow that they may be necessarily related; another is to deny that an explanation of gappiness entails a distinction in the properties involved in the first place. I'll begin with the first strategy.

Consider, then, the following argument. One can accept the inference from CP′ to the non-identity of qualia with any physical or functional property, and still reject PP′, so long as one insists that the relation between physical properties and qualia is metaphysically necessary. The point is that distinctness isn't sufficient; qualia must be only contingently related to their physical or functional correlates to get the metaphysical possibility of zombies out of their thick conceivability. But now the challenge is to make sense of the claim that though qualia are distinct from their physical correlates, they are metaphysically necessitated by them nonetheless.

One way is this. Drop the Modal Bridge Principle (MBP) articulated in section 1. MBP was introduced to solve an epistemological problem. By connecting metaphysical possibility and necessity to formal consistency and validity we could explain how we are able to determine not just what is actually the case, but what has to be the case. However, one could adopt the view that metaphysical necessity outstrips formal validity. There are metaphysically impossible situations that are not describable by any formally inconsistent description. Qualia are not identical to physical properties, nor is there any formally inconsistent description of the zombie situation, yet zombies are just not possible. Put another way, on this view the physical facts necessitate the mental facts by virtue of brute metaphysical necessity.[28]

I am not prepared to proclaim such a position incoherent, and therefore leave it as an option to be explored. But I find it quite implausible. For one thing, it either makes our epistemic access to modal facts quite mysterious, or it makes the modal facts epistemically inaccessible (those that outstrip the modal facts determined by logical form, that is). But more to the point, and this is connected to the epistemological problem, I'm not sure that the view really is coherent either. I just don't see what it could mean to judge impossible a situation that survives every test for logical consistency.[29] Of course I understand what it would mean to judge it nomologically impossible. But that is not sufficient for the materialist.

Brute metaphysical necessity seems to be a non-starter. Identity can do the metaphysical work needed to be done, but we've tentatively allowed that qualia are distinct properties. Mere nomological connection to physical properties isn't strong enough. Is there another relation that can do metaphysical work? Yes, there is a relation that is stronger than nomological connection but weaker than identity, and that is realization (or constitution, which is a member of the same family). If physical properties realize qualia, then it does follow that zombies are metaphysically impossible. So suppose we try that option.

But there is a serious problem with this solution as well. Essentially, for realization to do the requisite metaphysical work, it must rely on identity at a crucial point. Let me explain. Assuming MBP, as we are, the claim that zombies are metaphysically impossible entails the existence of a formal derivation from a description of the fundamental physical facts to some description of the qualitative facts. If we let "Bx" be the physical description, and "Rx" the mental one, we face the problem of disparate vocabularies. How do we get from a physical description to one that uses terms like "reddish"?

We might try this. Add "B realizes R" as a premise to the argument, much as we did with the identity claim "B = R" above. While it's true that from "Bx" and "B realizes R" one may infer "Rx", this just displaces the problem to the added premise. Unlike identity, which doesn't require explanation (assuming it's not gappy, that is), realization is a relation that admits of explanation. Not only does it make perfect sense to ask how it is that B realizes R, it is a question that demands an answer. To posit property realizations as a kind of brute fact, inex-

plicable in terms of other facts, is just as problematic as positing brute metaphysical necessities (perhaps more so).

What's required to explain realization is, to use Poland's (1994) term, a "realization theory".[30] A realization theory will show how it is that instantiating property B constitutes, or necessitates, instantiating property R. To demonstrate how B realizes R, then, seems to involve showing how we can derive a description of the facts represented in terms of "R" from a description of the facts represented in terms of "B". But now we're back to the original problem: deriving the "R-facts" from the "B-facts". The only way out seems to be to come up with a redescription of the R-facts in terms which allow a derivation from the appropriate description of the B-facts. Thus, we need an identity claim, of the form "R = X", for some appropriate "X", to ground the realization theory.

So it looks as if materialism is committed to some sort of identity theory. Maybe reddishness is not identical to a strictly physical, or even a biological property, but it must be identical to—that is, it must be—a property that admits of a description susceptible to derivation from physical descriptions. Our problem is that the only values for "X" that we can imagine substituting into "R = X", if we accept the argument of section 4, yield gappy identities. So long as we maintain that gappy identities entail a distinction in the relevant properties, we seem to be in trouble.

One final move seems available, though it is problematic. The materialist can hold out for there being a value of "X" to substitute into "R = X" above that yields a non-gappy identity. Perhaps more philosophical reflection, perhaps more empirical investigation, maybe a combination of both, will yield a conceptual breakthrough that will provide a genuinely satisfying realization theory for qualia.[31] The puzzle this move generates is obvious: what sort of breakthrough could the advocate of this position have in mind? The point is, we know a lot already about how the physical world is put together, and how information can be processed by physical systems. If the conceptual tools this knowledge provides aren't enough to bridge the explanatory gap, and therefore yield identity statements that aren't gappy, it's totally unclear what else could be on the horizon. Still, having no clue what the appropriate value for "X" could be is not the same thing as knowing definitively that there isn't one. For someone like myself who believes consciousness is a deeply puzzling phenomenon, but yet must be physical, this allows some room to maneuver.

Now let's consider the second strategy mentioned above. Let's challenge the assumption that to account for gappy identities we must appeal to a metaphysical distinction in properties. It's not that the assumption lacks plausibility; quite the contrary. In cases of non-gappy identities, such as "water = H_2O", while there is no a priori route from the "H_2O"-described facts to the "water"-described facts, still there is the definite sense that when all the chemical facts are in, the whole story's been told. As we argued above, no sensible question about how H_2O could be water remains. Thus the claim that the mere formal consistency of denying "water = H_2O" doesn't entail that there is a genuine distinction between the

properties involved seems fairly easy to accept. However, with the proposed identification of reddishness with a physical or functional property, where a substantive question does remain, the temptation to believe that there has to be some genuine distinction in properties corresponding to the representations of reddishness and its physical correlate is cognitively irresistible. Furthermore, we have the argument about extending our concept to new cases. The only way it seems plausible to distinguish cases where it seems to be a matter of semantics from those where it seems to be a matter of fact, is to claim that there are extra, metaphysically independent properties involved in the factual cases.

I find these considerations intuitively compelling, and I must admit I find it hard to see how qualia could actually be identical to either physical or functional properties. So that is one reason I take the first strategy seriously. Yet, despite these intuitively compelling considerations, the assumption they support—that gappy identities indicate distinct properties—does come into conflict with other considerations that are equally compelling. In particular, it seems to be based on a kind of Cartesian model of access to the facts; a model that is quite contrary to the spirit of MBP.

The point is, how can I tell, merely from facts about my own cognitive situation, including facts about various formal and semantic relations among my representations, that what one representation refers to is distinct from what another one refers to? The argument is supposed to be that only a distinction in the relevant properties could explain the gappiness of an identity. But gappiness is a matter of what I find intelligible, which in the end is a matter of how I represent the world. The bottom line is that my representations seem to present me with two distinct properties. But the possibility that distinct representations really refer to the same thing must always be an open one.

Suppose, then, we reject this crucial assumption underlying the inference from thick conceivability to possibility: that gappy identities reveal distinct properties. Now we are left with a different puzzle; namely, how to account for the distinction between gappy and non-gappy identities. The problem can be put this way. In both the gappy and non-gappy identities, we have two representations of the very same property, both of which involve non-ascriptive modes of presentation.[32] Non-gappiness is readily explained by the "behind-the-scenes" nature of non-ascriptive modes of presentation. If what we have in mind when we think of water is really just our mental representation of water, then we would expect there to be nothing cognitively left over after all the chemical and contextual facts were in. But gappiness is really puzzling. Causal or nomic relations don't seem apt for explaining the sort of cognitive relation we have to qualitative character. On the other hand, causal or nomic relations seem to be all the materialist has available to account for the representation relation.

I've considered two lines of reply to the anti-materialist thick conceivability argument. The first line allows the assumption that gappy identities reveal a metaphysical distinction in properties, and the second one denies it. Though both lines generate puzzles, I favor the second one for two reasons. First, I do think the

inference from gappy identity to a distinction in properties, though clearly quite compelling, is nevertheless fallacious for the reason cited above. Of course we seem to be wondering about a genuinely factual issue when we wonder whether some sufficiently similar creature is experiencing reddishly, even given all the facts about its physical and functional structure, and therefore what we're wondering about must involve the instantiation of a property distinct from the ones we already know about. Still, our conviction that this is so can't be a guarantee that it is.

But perhaps even more important—and this is a large topic I can only mention here—the puzzle we're left with when we challenge the assumption is one that we already have: how to account for there being a subjective point of view. What the gappiness of psycho-physical identities demonstrates is the existence of a kind of cognitive relation that is unique to our awareness of our own experiences. Whether or not qualia are distinct from their physical and functional correlates, it is still a difficult problem to explain the peculiar nature of first-person access to qualia. If we could explain how physical processes in the brain could give rise to this relation, we would go a long way to showing why there is an explanatory gap; and that would probably be the clue to actually bridging it.

Conclusion.

I presented two versions of the anti-materialist conceivability argument: one based on thin conceivability and the other on thick conceivability. The argument based on thin conceivability was rejected because it relied on an unwarranted assumption—that most concepts involve ascriptive modes of presentation, through which a priori, analytic connections to other concepts mediate cognitive access to the properties they represent. I argued that, on the contrary, there is no reason to suppose that concepts of natural kinds, in particular, involved such connections. So not only is a zombie thinly conceivable, but so is H_2O that isn't water. Thus thin conceivability is not a guarantee of possibility.

On the other hand, we saw that the argument from thick conceivability was much stronger. However, here too there is room for response. For one thing, the initial thick conceivability premise, to the extent that it entails much more than the corresponding thin conceivability premise, is more tendentious and less likely to be granted by materialists. Furthermore, even if it is granted, there are possible lines of reply, though they do generate puzzles of their own.

I would like to conclude with three observations. First, it might be thought that refuting the argument from thin conceivability is only relevant to one who rejects the thick conceivability premise but not the thin one. So long as one goes along with the thick conceivability of zombies, it might be thought that the entire argument concerning non-ascriptive modes of presentation is beside the point. However, that's far from true. Admittedly, it is difficult to make the thick conceivability of zombies intelligibly consistent with either the identity of qualia with physical properties, or their realization by physical properties. Still, there is more room to

maneuver than there would be if we were stuck with the assumption that whatever properties qualia are necessarily related to must be accessible a priori. It is much more plausible that we are just missing some crucial conceptual or theoretical insight which, if we had it, would render zombies no longer thickly conceivable, if the insight in question is not literally constitutive of our concept of a quale.

Second, the fact that there isn't a satisfying resolution of the thick conceivability of zombies with materialist metaphysics shouldn't be surprising. Thick conceivability is just another way of expressing the explanatory gap. Zombies are thickly conceivable because we don't have an explanation of qualia in terms of physical processes. Well, if we don't yet understand how qualia could be physical processes, then there ought to be something puzzling about the supposition that they are physical processes nevertheless. So long as the two crucial claims—that zombies are now thickly conceivable and that qualia are physical properties—are not literally inconsistent, materialism has a foothold even on the assumption that there is an explanatory gap. Again, were the mere thin conceivability of zombies—their conceptual possibility relative to the standard description—sufficient to undermine materialism, then it would be hard to see any possibility for a future vindication of materialism. But as things stand, we may not see yet quite how materialism could be true, but this doesn't commit us to its falsehood.

Finally, given how far I've been willing to come along the anti-materialist's path, it's a fair question to ask why not go the whole way. Perhaps the existence of the explanatory gap doesn't literally contradict materialism, there is still, as some might say, quite a bit of "tension" between them. But the problem is that whichever way you turn there are puzzles. For one thing, the problem of explaining our cognitive relation to experiential qualities, the problem of the subjectivity of experience, doesn't go away on the assumption that there are non-physical properties. More positively on the side of materialism, there is the problem of mental-physical causal interaction. I can't see any way to make sense of a qualitative property's causal relevance to other mental states and behavior unless it's either identical to, or realized in, physical properties, and I'm not prepared to give up on causal relevance.[33] Thus I am prepared to maintain that materialism must be true, though for the life of me I don't see how. In other words, the mind/body problem is just that: a problem.[34]

Notes

1. A word about "accuracy" is in order here. Consider the situation constituted by Bill Clinton's instantiating the property of not being President (in 1998, say—I'll leave this off in what follows). Clearly it is accurately represented by the sentence, "Bill Clinton is not President". But is it also accurately represented by "The President is not President"? If so, this appears to be a counterexample to MBP, since the latter sentence appears to be formally inconsistent, yet the situation is obviously metaphysically possible.

My response is that the sentence has two readings: on the first reading, it does accurately represent the situation, but it isn't formally inconsistent; on the second, though it is inconsistent, it doesn't accurately represent the situation. The first reading is the referential reading of "the Presi-

dent", as if it had a "dthat" operator attached. Given the logic of such expressions, it is not a logically inconsistent form in the relevant sense. The second reading is the general reading, on which it picks out the unique President in each possible world. But on that reading it doesn't accurately describe the situation in question, since in any world containing that situation—Bill Clinton not being President—Clinton doesn't satisfy the definite description, "the President". That is, Clinton's not being President is not a situation in which the unique President is not President.

I thank John Carroll, Randy Carter, and the Noûs referees for bringing up the objection that prompted this note.

2. In fact, it's clear that formal consistency couldn't be a more basic notion than metaphysical possibility, since there are an indefinite number of formal systems by reference to which one could define a notion of consistency. What makes the formal system we call "logic" the right one is the fact that its rules are truth-preserving, but the notion of truth-preservation is a modal one. The hypothesis represented by MBP is that our thought processes embody (roughly) the right formal system, the one that actually is truth-preserving, and it is this fact about us that explains our epistemic access to modal facts. See Rey (1993) for an interesting discussion of the idea that a priori knowledge of logic could be attributed to human beings on the basis of empirical considerations. My treatment of MBP seems to me consonant with Rey's view.

3. At this point let me enter a caveat. It doesn't appear as if the current framework can deal with mathematical necessity. After all, assuming that mathematical truth outruns logical truth (by just how much is controversial still, see Boolos 1975 and 1995), there will be mathematical truths that are not represented by logically valid forms. Since mathematical truths are metaphysically necessary, they constitute a counterexample to MBP. Of course I don't know that there is any truly adequate account of mathematical necessity, and I don't see that the application of this framework to the mind-body problem will be affected anyway. However, let me indicate at least one way one might try to deal with the problem.

One might bite the bullet and deny that mathematical truths, or those that aren't derivable from logic, are metaphysically necessary. While this may seem outlandish to some, I don't find it obviously crazy. If mathematics really isn't reducible to logic, then why not take the position that the mathematically possible worlds are a proper subset of the metaphysically possible ones? This is to say that the mathematical domain imposes substantive constraints on the way the world is, constraints that go beyond logic.

I'm not committed to this response, though I lean toward it. In general, as I said above, I think the case of mathematics, involving, as it does, a domain of abstract objects, is sufficiently different from the cases under discussion here that it can be set aside.

4. See Fodor (1975) and (1981, chapter 7) for extended defense of the so-called "language of thought" hypothesis.

5. Keep in mind here the remarks in note 1. The point is that the situation in which Aristotle is not a philosopher is not accurately described by the sentence "The great Greek philosopher who...is not a philosopher"; at least not on the general reading on which it's logically inconsistent.

6. I am indebted in my thinking about the relations among these various notions of possibility to the discussion in Yablo (1993). It's hard for me to say precisely how my characterization here relates to his. However, to the extent that we both see conceivability as a defeasible guide to possibility, there is a significant similarity.

7. In what follows I will drop any reference to "epistemic possibility". I will use "conceivable" in its narrower sense, in which it means (at least until section 4) "conceptually possible". We can think of the wider sense, according to which water's not being H_2O is now inconceivable, as a matter of all-things-considered conceivability. However, this wider sense is not going to play any significant role in our discussion. It was useful only as a means to introduce the narrower sense.

8. Though of course both Quine and his followers have denied a priori status even to logic as well. See Quine (1953) and Devitt (1996).

9. Of course some philosophers, such as Tye (1995) and Dretske (1995), are externalist about qualia too. I think there are real problems with this view, as I argue in Levine (1997). But nothing of

substance in this paper turns on denying externalism, since all the supervenience claims discussed could be modified to take account of relations to external objects.

10. From Smart (1959) to all those Block (1980) refers to as "psycho-functionalists" there is a long tradition in materialist writing that sees materialism as essentially an empirical hypothesis. Loar (1990) also explicitly rejects the a priori derivation response. Among those who deny the conceivability premise itself are Levin (1983), Lewis (1983a) and (1983b), and Shoemaker (1984, especially chapter 9).

11. One can see this reply as an assimilation of Smart's (1959) replies to his first two objections with the Kripke-Putnam account of natural kind terms.

12. This line of argument can perhaps be read back into Descartes, but has its contemporary source in Smart (1959); specifically his Objection 3. For more recent versions, see Chalmers (1996), Jackson (1980) and (1993), Kripke (1980), and White (1986).

13. A third option, which could be seen as related to this one, will be explored at length below.

14. This commitment can be seen in three ways. First, the argument for the DPM involved treating modes of presentation as "meanings", or analyses of what one has to know to count as competent with a term. That seems to entail that the relevant knowledge is a priori, based on analytic connections. Second, in order to avoid the consequence that the cup's being full of water is not metaphysically determined by the micro-physical facts, which would follow from a parallel conceivability argument, the anti-materialist must deny even the conceptual possibility of the relevant situation. But that means that the "water-facts" are derivable a priori from the micro-physical facts (together with some contextual/indexical information). Finally, both Chalmers (1996) and Jackson (1993) explicitly endorse the claim that there is an a priori derivation of the "water-facts" from the micro-physical facts.

15. For accounts along these lines see Fodor (1987) and Dretske (1981), among others.

16. See Fodor and Lepore (1992), also Antony (1993) and Levine (1993a).

17. The RDPM is the third option mentioned in note 11. The option we rejected above explained the conceptual possibility of "water is not H_2O" by appeal to the bare vocabulary difference. What was lacking was an account of competence with the vocabulary items that would still block a derivation from the one to the other. The RDPM does the job.

18. One can find versions of this argument in Bealer (1987) and Sidelle (1989).

19. One final point on this argument. It might be thought that the comparison to syntax actually supports the other side. After all, what the linguist infers from the capacity to sort sentences into the grammatical and ungrammatical (or, better, acceptable and unacceptable) is the existence of an underlying competence that consists in a representation of the rules that determine these judgments. So it looks as if one does infer from the capacity to form such stable intuitive judgments to the existence of explicitly represented rules, at least unconsciously.

There are two replies to this objection. First, not everyone who takes linguistics seriously feels that the rules of grammar must be explicitly represented. There is a raging controversy over this, and the question is a quite subtle one (see Stabler 1983). However, I don't want to rely on this reply since I am inclined toward the side that takes grammar to be explicitly represented.

So the more important point is this. If there is a good inference to an explicitly represented grammar that underlies grammaticality judgments, it isn't merely based on the existence of the capacity to make such judgments. It's a matter of inference to the best explanation. There is no analysis of what it is to have a grammatical capacity which entails the existence of an explicitly represented grammar. But the DPM advocate is arguing from an analysis of what it is to possess a concept to the claim that one must have a priori knowledge of how extension is determined by context. It is this inference that is being rejected. Whether implicit, or unconscious knowledge of the causal theory of reference is in fact the best explanation of our ability to render judgments in the Twin Earth case is an open question. But having that knowledge is not constitutive of having the concepts the causal theory is a theory of.

20. That is, I'm not imagining a situation where some paradigmatic sample of water is H_2O and much of the stuff in lakes and oceans is XYZ. While it's plausible that in such a case we would decide

that there are two kinds of water, as we have decided that there are two kinds of jade, I can see how the case might be filled out in such a way that we wouldn't say that, but rather that only H_2O is water.

21. I've heard Block use this example in many presentations, and it is mentioned in Block and Stalnaker (forthcoming).

22. This presumes that explanations involve deductions. I've defended retaining this feature of the DN model of explanation in Levine (1983) and (1993b).

23. For an argument along similar lines, see Block and Stalnaker (forthcoming).

24. For presentation of the argument that there is an explanatory gap, see the works cited in note 22.

25. Of course some would prefer to have "B" here refer to functional, as opposed to neurophysiological properties. Nothing of substance here will turn on this.

26. See Perry (1979).

27. Sydney Shoemaker, in his (1984), chapters 9 and 15, attempts to show how a functionalist theory could deal with this problem. I argue that his solution doesn't work in Levine (1989).

28. The position described here may be what Chalmers has in mind by "Strong Metaphysical Necessity", a position he rejects for reasons similar to those I give below. However, I'm not sure we are talking about the same thing.

29. Of course here again mathematical necessity rears its ugly head. Isn't the denial of the continuum hypothesis logically consistent, yet isn't it necessarily true if true at all? As I said earlier, I can't deal with the issue of mathematical necessity here. It seems obvious to me that whatever we end up saying about mathematics—and one option I take seriously is the view that the set of mathematically possible worlds is actually a proper subset of the metaphysically possible worlds—it won't be appropriate as a model for the metaphysics of mind.

30. A similar idea is called a "property theory" in Cummins (1983).

31. One might see this as actually a rejection of CP′, rather than PP′, since if we are in possession of the crucial non-gappy identity, zombies will no longer be thickly conceivable. Our realization theory, bridging the explanatory gap, will render zombies no different from zombie-H_2O. On the other hand, if you read the phrase "we can" in the definition of "thickly conceivable"—a situation is thickly conceivable relative to R just in case it's conceptually possible relative to R, and any derivation *we can* construct from R to a formally inconsistent representation R′, will include gappy identities in its premises—as restricting the range of derivations to be considered to those involving concepts available to us right now, then the move under consideration is consistent with CP′. I don't see that anything turns on which way we view it. If we take the first way, then we can say that for all we know zombies are thickly conceivable, but future conceptual progress may show us wrong.

32. I'm assuming this for simplicity. Of course one straightforward account of a gappy identity is that what we're seeking to explain is how the very same entity could share the distinct properties which are attributed to it through the distinct, ascriptive modes of presentation expressed by the terms flanking the identity sign. But our question is what to do when that move is unavailable; when we've come to the end of the line of appeals to properties ascribed in the mode of presentation.

33. For fuller discussion of this point, see Antony (1991) and Antony and Levine (1997).

34. An ancestor of this paper was written with support from a North Carolina State University College of Humanities and Social Sciences Summer Stipend, and the current version was completed while a Fellow of the American Council of Learned Societies. I thank both institutions for their support. Earlier versions were presented at UNC Chapel Hill (as a guest lecturer in Bill Lycan's NEH Summer Seminar), the University of Maryland (as a guest lecturer in Georges Rey's seminar), Virginia Technical Institute, and the Triangle Mind and Language Reading Group. I thank the audiences on those occasions for their helpful comments. Finally, I want to thank Louise Antony, David Auerbach, Ned Block, John Carroll, Randy Carter, David Chalmers, Doug Jesseph, Bill Lycan, Georges Rey, and Steve Yablo, and the Noûs referees for their comments, criticisms, and discussion of the issues in this paper.

References

Antony, L. (1991). "The Causal Relevance of the Mental: More on the Mattering of Minds," *Mind and Language*, vol. 6, no. 4.

——— (1993). "Conceptual Connection and the Observation/Theory Distinction", in Fodor, J. and Lepore, E., eds., *Holism: A Consumer Update*, special issue of *Grazer Philosophische Studien*, vol. 46.

Antony, L. and Levine, J. (1997). "Reduction With Autonomy", *Philosophical Perspectives, 11, Mind, Causation, and the World*, 83–105.

Bealer, G. (1987). "The Philosophical Limits of Scientific Essentialism", in J. Tomberlin, ed., *Philosophical Perspectives, I, Metaphysics*, Atascadero, California: Ridgeview Publishing Co.

Block, N. (1980). "Troubles with Functionalism", in N. Block, ed., *Readings in Philosophy of Psychology*, vol. 1, 268–305. Cambridge: Harvard University Press.

Block, N. and Stalnaker, R. (forthcoming). "Conceptual Analysis, Dualism, and the Explanatory Gap".

Boolos, G. (1975). "On Second-Order Logic", *Journal of Philosophy* 72 (16), 509–527.

——— (1995). "The Standard of Equality of Numbers", in *Frege's Philosophy of Mathematics*, W. Demopoulos, ed. Cambridge, MA: Harvard University Press.

Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.

Clark, Austen (1993). *Sensory Qualities*. Oxford: Oxford University Press.

Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge, MA: Bradford Books/The MIT Press.

Devitt, M. (1996). *Coming To Our Senses: A Naturalistic Program for Semantic Localism*. Cambridge: Cambridge University Press.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: Bradford Books/The MIT Press.

——— (1995). *Naturalizing the Mind*. Cambridge, MA: Bradford Books/The MIT Press.

Fodor, J.A. (1975). *The Language of Thought*. New York: Thomas Crowell, and Cambridge, MA: Harvard University Press.

——— (1981). *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: Bradford Books/The MIT Press.

——— (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: Bradford Books/The MIT Press.

Fodor, J. and Lepore, E. (1992). *Holism: A Shopper's Guide*. Blackwell: Oxford.

Hardin, C.L. (1988). *Color for Philosophers: Unweaving the Rainbow*. Indianapolis/Cambridge: Hackett Publishing Company.

Jackson, F. (1980). "A Note on Physicalism and Heat", *Australasian Journal of Philosophy*, vol. 58, no. 1.

——— (1993). "Armchair Metaphysics", in *Philosophy In Mind*, O'Leary-Hawthorne & Michael, eds., Dordrecht: Kluwer.

Kaplan, D. (1979). "DTHAT", in *Contemporary Perspectives in the Philosophy of Language*, French, Uehling Jr., & Wettstein eds., University of Minnesota Press.

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Levin, J. (1983). "Functionalism and the Argument from Conceivability", *Canadian Journal of Philosophy*, Supplementary Volume 11.

Levine, J. (1983). "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly* 64.

——— (1989). "Absent and Inverted Qualia Revisited", *Mind & Language*, vol. 3, no. 4.

——— (1993a). "Intentional Chemistry", in Fodor, J. and Lepore, E., eds., *Holism: A Consumer Update*, special issue of *Grazer Philosophische Studien*, vol. 46.

——— (1993b). "On Leaving Out What It's Like", in Davies, M. and Humphreys, G., eds., *Consciousness: Psychological and Philosophical Essays*, Oxford: Blackwell.

——— (1995). "Qualia: Intrinsic, Relational, or What?", in *Conscious Experience*, Thomas Metzinger, ed., Sch[..]oningh Verlag/Imprint Academic.

——— (1997). "Are Qualia Just Representations? A Critical Notice of Michael Tye's *Ten Problems of Consciousness*", *Mind and Language*, Vol. 12, No. 1, 101–113.

Lewis, D. (1983a). "Mad Pain and Martian Pain", *Philosophical Papers*, vol. 1. Oxford: Oxford University Press.

——— (1983b). "Postscript to 'Mad Pain and Martian Pain' ", *Philosophical Papers*, vol. 1. Oxford: Oxford University Press.

Loar, B. (1990). "Phenomenal States", *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, J. Tomberlin, ed., Atascadero, California: Ridgeview Publishing Co.

McGinn, C. (1991). *The Problem of Consciousness*. Oxford: Basil Blackwell.

Nagel, T. (1974). "What Is It Like to Be a Bat?", *The Philosophical Review*, vol. 82.

Perry, J. (1979). "The Problem of the Essential Indexical". *Nous* 13.

Poland, J. (1994). *Physicalism: The Philosophical Foundations*. Oxford: Clarendon Press.

Quine, W. V. (1953). "Two Dogmas of Empiricism", in *From a Logical Point of View*, Harvard University Press: Cambridge, MA.

Rey, G. (1993). "The Unavailability of What We Mean I: A Reply to Quine, in Fodor, J. and Lepore, E., eds., *Holism: A Consumer Update*, special issue of *Grazer Philosophische Studien*, vol. 46, 61–10.

Shoemaker, S. (1984). *Identity, Cause, and Mind*. Cambridge: Cambridge University Press.

Sidelle, A. (1989). *Necessity, Essence, and Individuation: A Defense of Conventionalism*. Ithaca: Cornell University Press.

Smart, J.J.C. (1959). "Sensations and Brain Processes". *The Philosophical Review*, 68.

Stabler, E.P. (1983). "How are Grammars Represented?", *The Behavioral and Brain Sciences*, vol. 6.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: Bradford Books/M.I.T. Press.

Van Gulick, R. (1993). "Understanding the Phenomenal Mind: Are We All Just Armadillos?", in Davies, M. and Humphreys, G., eds., *Consciousness: Psychological and Philosophical Essays*, Oxford: Blackwell, 138–154.

White, S. L. (1986). "Curse of the Qualia". *Synthese* 68.

Yablo, S. (1993). "Is Conceivability a Guide to Possibility?", *Philosophy and Phenomenological Research*, Vol. LIII, No. 1.