

To appear in M. O'Rourke and C. Washington (eds.), *Situating Semantics: Essays on the Philosophy of John Perry* (MIT Press)

A REFUTATION OF QUALIA-PHYSICALISM

Michael McKinsey
Wayne State University

Recent defenders of reductive physicalism such as Brian Loar (1990, 1997) and John Perry (2001) have adopted an intriguing new strategy:¹ (1) accept as so much common sense (nearly) everything that property-dualists want to say about sensory qualia, including the central claims that sensory qualia determine 'what it's like' to have a given sense experience, and that persons are directly aware of these qualia in the having of such experiences; (2) contend that while these common sense facts about qualia may show that our ways of thinking and speaking about qualia are *conceptually* different from our ways of thinking and speaking about physical properties of the brain, these facts do *not* show that sensory qualia *themselves* (as opposed to our ways of thinking and speaking about them) are distinct from physical properties of the brain; (3) use this contention to turn aside the few existing arguments against reductive physicalism by such property dualists as Kripke (1972), Nagel (1974), Jackson (1982), and Chalmers (1996); and finally (4) insist that sensory qualia are in fact just identical with physical properties of the brain, so that consequently, the facts about the sensory qualities of conscious experience are nothing over and above physical facts about the brain. I will call the view that incorporates this strategy 'qualia-physicalism', or 'Q-physicalism' for short.

In this paper I will argue that Q-physicalism is false. I agree that in the case of *some* words, notably indexical pronouns, proper names, and natural kind terms, there is an important semantic distinction between (a) the word's conceptual or linguistic meaning, and (b) the particular contribution, which I call 'propositional meaning', that the word makes to the propositions expressed by use of sentences containing the word. This distinction makes it possible

in principle for there to be predicates that have distinct conceptual meanings and that nevertheless ascribe or predicate the same properties. However, or so I shall argue, the conceptual facts about predicates that ascribe sensory qualia, on the one hand, and predicates that ascribe physical properties to the brain, on the other, are sufficient to guarantee that the logical properties of what we say, the propositions we express, when we ascribe sensory qualia to our experiences, are quite different from the logical properties of what we say, the propositions we express, when we ascribe physical properties to the brain. It follows that propositions about the sensory qualities of conscious experience, and hence the facts that make these propositions true, are *not* the same as any physical propositions and facts about the brain. Hence, Q-physicalism is false.

The question of whether Q-physicalism is true or false is crucial to our understanding of the mind-body problem. This is because the phenomenal qualities of sense experience are so difficult to incorporate in a physicalist world view, and it seems to me that Q-physicalism is the only form of *reductive* physicalism that is capable of taking qualia seriously. The Q-physicalists, in contrast for example to defenders of functionalist views of mental properties, can agree with non-reductivists like myself that sense experiences are distinguished by their qualitative features, features that determine ‘what it’s like’ to have the experience in question.

The Q-physicalist’s distinctive claim is of course that these qualitative features of sensations are identical with physical properties of the brain, and it is this claim that I’m most interested in refuting.ⁱⁱ If I am right, then sensory qualia are not identical with any physical properties of the brain. Sensory qualia, I take it, are also not functional properties that logically supervene upon physical properties, since no functional property is sufficient to determine ‘what it’s like’ to have any given sense experience.ⁱⁱⁱ Thus, if Q-physicalism is false, then it surely seems that sensory qualia are not reducible in any way to physical properties, and so no form of

reductive physicalism is true: the phenomenal properties of sense experience, and the propositions and facts involving the ascription of such properties, are properties, propositions, and facts ‘over and above’, and are not reducible to, any physical properties, propositions, or facts. So if I am right, only non-reductive forms of physicalism have a chance of being true.

My arguments against Q-physicalism will rely crucially on claims about logical possibility, and in this respect they will resemble some of David Chalmers’ (1996) recent important arguments against reductive physicalism. Certainly, my discussion will be very much akin to Chalmers’, both in its spirit and in its conclusion. However, as we shall see, the concept of logical possibility on which I will rely is significantly different from the concept that Chalmers uses, making my arguments quite different from his, and allowing me to avoid certain objectionable assumptions about possible worlds that Chalmers’ arguments require. I will discuss Chalmers’ concept of logical possibility in detail below, after I’ve given my arguments against Q-physicalism.

1. Descartes’ argument for substance-dualism.

In order to motivate and explain the concept of logical possibility on which I will be relying, it will be useful to consider an argument that I think may capture Descartes’ chief reason for believing that he is not a physical object.^{iv} To state the argument succinctly, let ‘**d**’ be a name of Descartes, and define the property **P** as follows:

x has **P** =_{df} it is possible that: **x** exists, and there are no physical objects.

Then the argument is simply:

- (D) 1. **d** has *P*.
 2. No physical object has *P*.
 ∴ 3. **d** is not a physical object.

This argument certainly has a valid form. Moreover, each of its premises, taken separately, is plausible. Premise (D1) is plausible because as Descartes argued at length in *Meditation II*, it is consistent with all he knows for certain about himself as a thinking thing that he is a disembodied mind in a purely spiritual universe. So it certainly does not *logically follow* from the assumption that Descartes exists, that there are any physical objects. In this sense, then, it is possible that Descartes exists in a world devoid of physical objects, and premise (D1) is true.

Premise (D2) is also plausible, but for a quite different sort of reason. Any physical object, it seems, would have to have a physical *nature*, or essence; such an essence would be a physical property without which the object in question could not exist. Thus no physical object could exist in a possible world in which it lacks its physical essence. Hence no physical object could exist in a possible world devoid of physical objects, for the simple reason that no physical object could exist in a possible world in which it itself is not a physical object. So in this sense, premise (D2) is true.

But by now the reader should have begun to suspect that Descartes' argument equivocates on two senses of the word 'possible'. Premise (D1) is true when 'possible' is understood in a very weak sense to mean 'logically possible'. In this weak sense, a proposition is logically possible just in case it is logically *consistent*, that is, fails to logically imply a contradiction. Let us say that a proposition *p* *logically implies* a proposition *q* just in case there is a correct deduction of *q* from *p*, a deduction whose only premises other than *p* are either logical truths or other necessary, conceptual truths that are knowable *a priori*, and every step of which immediately follows from

previous steps by a valid rule of inference.^v Then we can say that it is *logically possible* that p if and only if the proposition that p fails to logically imply any explicit contradiction (of the form $(q \ \& \ \sim q)$). Note also that it follows from these definitions that if a proposition p fails to logically imply a proposition q , then it is logically possible that $(p \ \& \ \sim q)$.^{vi} (Note that these senses of logical implication and logical possibility are *broadly* rather than strictly logical, since I assume that these senses can apply or fail to apply by virtue of necessary *a priori* conceptual truths that may not be strict logical truths.)

As we've seen, premise (D1) is true, given that the property P is understood in terms of mere logical possibility. But premise (D2) is *false* when 'possible' is understood in this weak sense. For just as the existence of Descartes fails to logically imply that there are any physical objects, so the mere existence of *any* object x fails to logically imply that there are physical objects, *whatever* that object x 's actual nature happens to be. Consider for instance the large, battered, old oak desk at which I am now writing, and call it 'Fred'. Fred is of course the very paradigm of a physical object. But the simple singular proposition about Fred to the effect that Fred exists does *not* logically imply that there are any physical objects: no logical or conceptual truths that are knowable *a priori* will allow us to prove that there are physical objects, given merely the assumption that Fred exists.

Of course we know that Fred has a physical nature, since Fred is a large oak desk. Even so, it no more *follows* from Fred's existence that there are physical objects, than it follows from Descartes' existence that there are such things. In fact, in my view, Descartes and Fred are in exactly the same boat. For just like Fred, Descartes also has a physical nature. It's just that in the case of neither object does it follow from the object's mere existence that the object has a physical nature.

The sense of ‘possible’ in which premise (D2) is true is much stronger than mere logical possibility. I will call this stronger sense ‘metaphysical possibility’. A proposition *p* is *metaphysically possible* just in case there is a (metaphysically) possible world *w* such that *p* is true in *w*. Since every physical object has a physical essence, no physical object can exist in a possible world in which there are no physical objects. So premise (D2) is true in this sense: for no physical object is it metaphysically possible that that very object exists, even though there are no physical objects.

But taken in this sense of ‘possible’ premise (D1) is at best highly controversial, and Descartes has given us no reason whatever to believe that it is true. For Descartes has provided support for only the weak claim that his existence does not logically imply that there are physical objects. And as Father Arnauld pointed out, the fact that Descartes could not *prove* that he has a physical nature from premises about himself that he knew with certainty, hardly shows that Descartes *does not in fact* have a physical nature.^{vii}

Arnauld was justifiably very puzzled as to why Descartes should have thought otherwise. The solution to the puzzle would seem to be that, in discovering that his existence as a thinking thing failed to logically imply that there are physical objects, Descartes assumed that he had also discovered a sense in which it is *possible* for him to exist as a disembodied mind in a purely spiritual universe. But what Descartes failed to notice is that this was a *new* sense of possibility, a sense quite different from the then traditional, metaphysical sense that Arnauld had in mind. In this traditional sense (no doubt inherited from Aristotle) what is possible is not determined by logic alone, but also by facts about the *natures* of things, where some of these facts can only be known *a posteriori*.

Descartes' discovery of a new kind of possibility (and a corresponding new kind of necessity) was certainly an historically important event, since it seems to have eventuated in the Humean idea that the necessary truths are just the *analytic* truths, an idea that has been so dominant in the recent past. But unfortunately, Descartes immediately *confused* his important new sense of possibility with the old sense: for again, as Arnauld pointed out, nothing about Descartes' nature follows from the mere logical possibility that Descartes exists in a world devoid of physical objects.

I intend the distinction between metaphysical and logical possibility, as I've explained these notions, to apply only to *propositions*, rather than to *worlds*.^{viii} Earlier I used the notion of a possible world to explain a proposition's being metaphysically possible as meaning that the proposition is true in some (metaphysically) possible world. Thus any metaphysically possible proposition must also be logically possible, since of course no logically impossible (inconsistent) proposition is true in any possible world. But the converse does not hold: a proposition, such as the proposition that Descartes is not a physical object, can be both logically possible and metaphysically *impossible*.

This might suggest that a logically possible proposition is one that is true in some *logically* possible world, and that therefore, the metaphysically possible worlds are simply a proper subclass of the logically possible worlds. But like many others, I find this suggestion to be counterintuitive.^{ix} For instance, given that water is necessarily composed of H₂O molecules, there is no metaphysically possible world in which water is not H₂O. But according to my definition, it is certainly logically possible that water is not H₂O. Should we then say that in some *logically* possible world, water is not H₂O? Apparently not. Since water cannot exist without being composed of H₂O molecules, it seems wrong to suppose that in *any* member of *any* subclass of

the possible worlds, water exists and yet is not composed of H₂O. Thus the concept of logical possibility that I'm using requires me to *reject* any inference from a premise of the form 'it is logically possible that *p*' to a conclusion of the form 'for some possible world *w*, it is true that *p* in *w*'. (In this respect my concept of logical possibility contrasts sharply with that of Chalmers, which I will discuss at length below in sections 10 and 11.) For me then, the class of possible worlds = the class of metaphysically possible worlds, and thus I *accept* inferences from premises of the form 'for some possible world *w*, it is true that *p* in *w*' to conclusions of the form 'it is metaphysically possible that *p*', and *vice versa*.^x

2. A Cartesian argument against Q-physicalism.

We have seen that Descartes' argument from logical possibility fails to support *substance-dualism*, the view that persons are not physical objects. Nevertheless, considerations regarding logical possibility of the sort that Descartes used to argue against substance-dualism do suffice, I believe, to show that *type-dualism* is true, where type-dualism is the view that many mental properties, propositions, and facts are distinct from any physical properties, propositions, and facts. In particular, considerations of the sort that Descartes adduced are sufficient to show that sensory qualities of experiences are not physical properties of any sort and hence that Q-physicalism is false.

Let us say that a property is a *physical property* just in case possession of that property by a concrete entity such as an object, event, or state logically implies that the object, event, or state in question is itself a physical entity, that is, an entity that has a location in physical space.^{xi} As I am using the term, *propositions* are the abstract semantic contents, or meanings, expressed by the declarative sentences of natural languages. Propositions are what such sentences can be used to *say* or *assert*, and they are also the contents of cognitive attitudes, such as *thoughts* and *beliefs*.

Propositions are also, I assume, the ultimate bearers of truth-value: the truth or falsity of a proposition p determines the truth or falsehood of any sentence which expresses p , and of any thought or belief that has p as its content. By a *physical proposition*, I will mean a proposition that logically implies the existence of physical entities. Thus any proposition that ascribes a physical property to an object, event, or state will be a physical proposition. By a *physical fact*, I will mean a fact that corresponds to, or is the objective correlate of, a true physical proposition.^{xii} As I use the term, facts are precisely as intensional and fine-grained as the true propositions to which they correspond. Thus for me, the fact that p = the fact that q only if the proposition that p = the proposition that q .

As I say, Descartes' discussion in the *Meditations* concerning what he could consistently doubt given incorrigible knowledge of himself and many of his mental states, acts, and experiences, suffices to show that Q-physicalism is false. For Descartes' discussion shows that ever so many true propositions concerning one's mental states, acts, and experiences fail to logically imply the existence of one's own body, and indeed fail to logically imply the existence of even so much as a single physical object. Suppose for instance that a given person x feels a sharp pain. It certainly does not seem to *logically follow* from the assumption that x feels a sharp pain that x is a physical object. Hence it certainly *seems* that possession of the mental property of feeling a sharp pain by x does not logically imply that x is a physical object, and if so, then this mental property is simply *not* a physical property, and in particular, it is not a physical property of the brain. Hence Q-physicalism is false.

3. The Q-physicalist's reply.

Defenders of Q-physicalism will no doubt fail to be impressed by this simple, not to say simple-minded, argument. They would perhaps immediately respond by accusing the argument of 'begging the question' against their view. After all, on their view, the phenomenal properties that we truly ascribe to ourselves and to our mental experiences just *are* physical properties of the brain. So the true propositions asserted by such ascriptions, propositions about ourselves and our mental experiences - our feelings of pain, for example - just *are* physical propositions about our brains. So contrary to what Descartes thought, these propositions logically imply the existence of physical objects after all.

But how could Descartes (like the rest of us) have missed so many of the *logical* implications of the mental propositions which he knew incorrigibly and apparently, understood so well? The Q-physicalist is ready with an answer. Consider the forms of sentence that we normally use to describe the qualitative features of sensations, for example a form such as

(P) *X* feels a sharp pain.

According to the Q-physicalist, a sentence of this form would ascribe a property that is in fact a physical property of the brain. Apparently, the idea is that the brain in question would be the brain of the person whose sense experience is being described (see for instance Perry, 2001, pp. 36-7). So a sentence of the form (P) might ascribe the same property as some sentence written in neurophysiological language, such as

(B) The C-fibers of *X*'s brain are firing to degree *n*.

Now the Q-physicalist is quick to point out, indeed insist upon, the fact that sentence forms like (P) and (B) are radically different from each other in semantically important ways.

Thus it is obvious, as everyone can agree, that (P) and (B) have quite different linguistic *meanings*. A language such as English could contain sentence forms with the meaning of (P) and yet completely lack expressions with the meaning of (B), and *vice versa*. Thus speakers of such a language could have perfect understanding of the meaning of a sentence form like (P), and yet fail completely to understand any sentence form with the meaning of (B), and *vice versa*.

In short, the Q-physicalist points out, the forms (P) and (B) are *conceptually* different: the concepts one must have to understand (P), such as the concepts of *feeling* and *pain* are admittedly quite different from the concepts required to understand (B), such as the concepts of *brain* and *C-fiber*. Nevertheless, this conceptual difference is consistent with the possibility that these forms express, and are used to predicate, one and the same *property*. And this is just what the Q-physicalists seem to be claiming: sentence forms like (P) and (B), despite their conceptual differences, can nevertheless be used to ascribe one and the same physical property. (See for instance Loar 1990, pp. 81-5. See also Perry 2001, Chapter 3, pp. 68-69. Perry could put this point by emphasizing that while such forms as (P) and (B) differ radically in what he calls their *reflexive* contents, they can nevertheless coincide in their *subject-matter* content, which in this case would be the common physical property represented by the forms.).

Now it certainly seems that any two sentences that ascribe the same properties to the same objects must express the same propositions as well as state the same facts. That this is so follows from a very plausible and commonly endorsed Fregean principle of compositionality for propositions, namely

(FRE) The proposition expressed by a given sentence is a *function* of, or is determined by, the properties ascribed by the predicates contained in the sentence.

It follows from **(FRE)** that if two sentences differ only in that the sentences contain distinct predicates, then if the predicates ascribe the same property, the sentences must express the same proposition. On the Q-physicalist hypothesis we're considering, the predicates contained in (P) and (B) ascribe the same property, and so instances of (P) and (B) concerning the same person must express the same proposition. It then follows that the propositions expressed by such instances of (P) and (B) must share all their logical properties in common.

In particular, the proposition that the C-fibers of one's brain are firing to degree *n* *logically implies* the existence of brains and C-fibers. But if instances of (P) ascribe the same property as instances of (B), then by our Fregean principle **(FRE)** such instances that concern the same object must express the same propositions, and hence the propositions expressed must have all the same logical properties. Hence, the proposition that one feels a sharp pain, being identical with the proposition that the C-fibers of one's brain are firing to degree *n*, must also logically imply the existence of brains and C-fibers, and so if Q-physicalism is true, then it is *not* logically possible after all that one can feel a sharp pain, even though there are no such things as brains and C-fibers.

Of course, as Descartes' discussion in the *Meditations* shows, it certainly does not *seem* to logically follow that there are such things as brains and C-fibers, given merely the premise that one is feeling a sharp pain. This appearance is due to the fact that one cannot construct an *a priori* deduction, one based on understanding of concepts and meanings alone, from the mere assumption that one is feeling a sharp pain to the conclusion that there are brains and C-fibers. However, the Q-physicalist contends, this failure of *a priori* deducibility is fully explained by the *conceptual* differences between such linguistic forms as (P) and (B). What Descartes' discussion shows is merely that it is *conceivable* that one might feel a sharp pain even though there are no

such things as brains and C-fibers. But again, this is not really even *logically* possible, given that the forms (P) and (B) do in fact ascribe the same property.

Notice that, if the Q-physicalist is right, then one's inability to construct an *a priori* deduction from the assumption that one feels a sharp pain to the conclusion that there are brains and C-fibers does *not* by itself show that the proposition assumed fails to *logically imply* this conclusion, in my sense of 'logically imply'. For given the Q-physicalist's thesis that the proposition that one is feeling a sharp pain is *identical* with the proposition that the C-fibers of one's brain are firing to degree *n*, it follows that there is *a* correct deduction from this proposition as premise to the conclusion that there are brains and C-fibers. It's just that in order for one to *know (a priori)* that this deduction is correct, one must *conceive* the premise *as* the proposition that the C-fibers of one's brain are firing to degree *n*. If instead one *conceives* the premise *as* the proposition that one is feeling a sharp pain, then one will be unable to tell (*a priori*) that this premise logically implies that there are brains and C-fibers.

I think that most philosophers, myself included, would agree that conceivability is an unreliable guide to what is *metaphysically* possible. (See for instance, Yablo 1993.) But we've just seen that Q-physicalists must apparently say in addition that conceivability is also an unreliable guide to what is *logically* possible. Now that conceivability does not guarantee logical possibility is far from being obviously correct, and as far as I know, no Q-physicalist has explicitly defended it. Yet as we've just seen, the Q-physicalist's distinction between the *concept* and the *property* expressed by a mental predicate does provide the basis of such a defense. Moreover, as we shall see, there is a grain of truth in this defense. For Chalmers' (1996) hypothesis that conceivability at least guarantees logical possibility is in fact wrong: as we shall see, there are cases in which it is conceivable that *p* even though it is *not* logically possible that *p*.

4. Two forms of Q-physicalism.

Although the defense I just stated on behalf of such Q-physicalists as Loar and Perry does closely resemble, and is certainly consistent with, what the Q-physicalists have actually said in print, I can't be certain that all or even any of the Q-physicalists would agree with the defense exactly as I've stated it. The crucial claim of the form of Q-physicalism just defended is that mental predicates such as that found in the form (P) are actually used to *predicate* or *ascribe* sensory qualities that turn out to be identical with physical properties of the brain, properties that would also be ascribed by such a complex neurophysiological predicate as that found in the form (B). This claim is required for the application of our Fregean principle (**FRE**), which then has the further consequence that the mental propositions and facts expressed and stated by instances of such forms as (P) turn out to be identical with physical propositions and facts, since they turn out to be identical with neurophysiological propositions and facts expressed and stated by instances of such forms as (B).

To have a name for this form of Q-physicalism, I will call it *predication of qualia*-physicalism, or PQ-physicalism for short. Although I think it is natural to take the Q-physicalists to be defending and endorsing PQ-physicalism, there is another view which they (or some of them) may have in mind instead. As far as I can tell, the Q-physicalists never speak of phenomenal mental predicates as *predicating* or *ascribing* physical properties. Rather, they pretty uniformly say that phenomenal concepts and predicates *refer* to (what turn out to be) complex physical properties of the brain. (See for instance Loar, 1990, p. 84; and Perry, 2001, p. 64.) So perhaps these Q-physicalists would not agree that any phenomenal mental predicates actually *predicate* neurophysiological properties, and would prefer to say instead that such predicates

merely *refer* to, or perhaps merely *involve* reference to, physical properties of the brain. I will call this view *reference to qualia*-physicalism, or RQ-physicalism for short.

Although I think that both PQ-physicalism and RQ-physicalism are false, we shall see below that they are different views with different implications, and so they are false for different reasons. In particular, PQ-physicalism entails, while RQ-physicalism does not, that the mental propositions expressed by instances of such forms as (P) are identical with physical propositions about the brain expressed by instances of such forms as (B). Thus PQ-physicalism entails, while RQ-physicalism apparently does not, that instances of such forms as (P) logically imply the existence of such things as brains and C-fibers. The version of Q-physicalism defended earlier was really PQ-physicalism. I will at this point continue my discussion of PQ-physicalism, argue that it is false, and then later I will turn to discussion of RQ-physicalism.

5. Two types of meaning.

Both forms of Q-physicalism that we've identified must rely on what is by now a well known distinction between two types of meaning that I call *linguistic* meaning and *propositional* meaning. Our current understanding of this distinction is due primarily to the seminal work on indexical and demonstrative pronouns by David Kaplan (1977, 1978, 1979) and John Perry (1977, 1979).^{xiii} I have elsewhere described and defended this distinction in some detail for the cases of indexical pronouns, proper names, and natural kind terms. (See McKinsey 1984, 1987, 1994, and 1999.) On my account, the linguistic meaning of a word is the meaning that the word has in a given language, by virtue of a semantic rule of the language that governs the word's use. For instance, the linguistic meaning of the indexical pronoun 'I' in English would seem to be provided by a rule of reference such as

- (I) For any token α of 'I' and any object x , α is to refer to x if and only if x = the speaker of α .

By contrast, the *propositional* meaning of a word is the specific semantic contribution that the word makes toward determining which proposition is expressed by a sentence containing the word. The distinction between linguistic and propositional meaning is easiest to see in the case of indexicals like 'I', since such pronouns can be used with a *single* linguistic meaning (provided by a rule like (I)), while that very meaning requires the pronoun's propositional contribution to *vary* from one occasion of use to another (as Frege (1956) first pointed out). For instance, a sentence like 'I am watering the lawn', though it has a single linguistic meaning, will express different propositions on different occasions of use, depending on the identity of the speaker and the time of utterance. Hence, distinct utterances of 'I' must be taken to make distinct propositional contributions in distinct utterances of an indexical sentence like 'I am watering the lawn'.

Since the proposition expressed (on an occasion) by a sentence containing an indexical pronoun seems to be determined by that indexical's *referent* (on that occasion), the most plausible view is that an indexical's propositional meaning (on an occasion) just *is* its referent (on that occasion). So in the case of indexicals, the distinction between linguistic and propositional meaning boils down to the distinction between (linguistic) meaning and *reference*. And since the linguistic meaning of an indexical pronoun is given by a rule like (I) that determines the indexical's referent (on an occasion), we have in this case the consequence that *linguistic meaning determines propositional meaning*.^{xiv} A term whose linguistic meaning determines that the term's propositional meaning (on an occasion) just *is* the term's referent (on that occasion) is what I call a *genuine term*.^{xv} The proposition expressed by a sentence containing a genuine term is a *function*

of the term's referent. Such propositions are standardly called *singular* propositions, and genuine terms are standardly said to *directly refer* to their referents.

So far, we've seen that a sentence containing an indexical, a sentence like 'I am watering the lawn', can have a single linguistic meaning even though it can be used to express many different propositions. But for our present purpose of evaluating PQ-physicalism, the more important semantic fact is that different indexical sentences, with different linguistic meanings, can be used to express the *same* proposition. For instance, suppose that a friend looking for my house calls me on his cell phone, saying "I'm in your neighborhood but I can't find your house. ...Wait, there's a guy at the end of the block.... *he's watering the lawn.*" I then reply (on my cell phone), "That's odd, *I'm watering the lawn* too." In fact of course I'm the man my friend sees at the end of the block, and so he and I have said the same thing, expressed the same proposition, with our different sentences ('he's watering the lawn' and 'I'm watering the lawn'). Moreover, while each of us has perfect understanding of the (linguistic) meaning of all the sentences being uttered, it may well be that, at least at first, neither of us realizes that we've asserted the same proposition. The reason why this can happen is that sometimes, the proposition one expresses is determined by the *referent* of a word one has used, even though the referent of that word is not determined *completely* by the word's linguistic meaning alone. For instance, suppose that the linguistic meaning of the demonstrative 'he' is given by the semantic rule

(He) For any token α of 'he' and any object x , α is to refer to x if and only if x = the male that the speaker of α demonstrates by uttering α .

Then in the case described, the fact that I'm the referent of my friend's utterance of 'he' is determined partly by the meaning of his utterance of the pronoun 'he' and partly by the (non-semantic) fact that I happen to be the man my friend sees down the block, so that as it turns out, I

am in fact the male my friend has demonstrated with his utterance of ‘he’. Being at first unaware of these non-semantic facts, I don’t at first realize that my friend has said that I’m watering the lawn.

This case involving indexical pronouns bears important similarities to what the PQ-physicalist wishes to say about the semantic relations between such sentence forms as (P) and (B) above. (Similarities of this sort are emphasized and described in various useful ways throughout the text of Perry, 2001.) Perhaps a person (Descartes, say) could have complete understanding of the linguistic meaning of an uttered instance of ‘*X* feels a sharp pain’ and yet still be unaware of certain important non-semantic facts that in part determine what proposition is expressed by that utterance. And perhaps these non-semantic facts, together with the linguistic meaning of (P), have the result that the person’s utterance of an instance of (P) expresses the same proposition as an instance of (B). In such a case, perhaps, the person has - though with no awareness of having done so - actually asserted a proposition that *logically implies* the existence of such things as brains and C-fibers. This could happen, if the *linguistic* meaning of the predicate ‘feels a sharp pain’ were such as to allow it to turn out that the predicate’s *propositional* meaning just is the property ascribed by instances of (B), namely, the property of having a brain whose C-fibers are firing to degree *n*. (Or, in Perry’s terminology, this could happen if the *reflexive* content expressed by the predicate were such as to allow it to turn out that the predicate’s *subject-matter* content just is the physical property ascribed by instances of (B). See Perry 2001, pp. 68-69.)

6. The case of proper names.

Perhaps, then, the distance between linguistic and propositional meaning allows for the possibility that the propositions we express with the words we use have *logical* properties of which we are ignorant, properties of which we can have no *a priori* knowledge, since the

conceptual meanings of the words we use can by themselves provide no basis for knowledge of the properties in question.

This possibility is even more strongly suggested by the case of *proper names*. The modal evidence that Kripke (1972) provided to show that every proper name is a rigid designator (a term that refers to the same object relative to every possible world) is also strong evidence that names, like indexicals, are genuine terms. If so, then two sentences which are just alike except for containing distinct names must express the same proposition if the two names have the same referent. But two such sentences that express the same proposition might well have different conceptual meanings, if the names themselves have different linguistic meanings.

The classic example of this is of course provided by identity sentences such as ‘Hesperus = Hesperus’ and ‘Hesperus = Phosphorus’ (‘H=H’ and ‘H=P’, for short). Given that the names ‘Hesperus’ and ‘Phosphorus’ are genuine terms that have the same referent, these two sentences, despite appearances, must express the same proposition. From the apparent difference in ‘cognitive value’ of such pairs of identity sentences, Frege (1892) reasonably inferred that the members of such pairs must express different propositions. But as we’ve just seen, one who holds as I do that names are genuine terms must give some other explanation of why ‘H=H’ and ‘H=P’ differ in cognitive value.

One possibility is that ‘Hesperus’ and ‘Phosphorus’, though both are genuine terms whose referent (the planet Venus) is their only propositional contribution, also have distinct *descriptive meanings* that determine, or fix, the names’ reference in Kripke’s (1972) sense.^{xvi} Perhaps, suppose, the linguistic meanings of ‘Hesperus’ and ‘Phosphorus’ are given by the following different descriptive reference rules:

(HES) For any token α of ‘Hesperus’ and any object x , α is to refer to x if and only if x = the heavenly body that appears brightest on the western horizon in the evening.

(PHOS) For any token α of ‘Phosphorus’ and any object x , α is to refer to x if and only if x = the heavenly body that appears brightest on the eastern horizon in the morning.

Given this hypothesis about the meanings of ‘Hesperus’ and ‘Phosphorus’, the sentences ‘H=H’ and ‘H=P’ will differ in linguistic meaning and in ‘cognitive value’, even though the sentences express the same proposition. But notice that on this same hypothesis, one cannot know *a priori* that the sentences *do* express the same proposition. Certainly, one can know, just on the basis of knowledge of meanings alone, that *if* the sentence ‘H=P’ is true, then it expresses the same proposition as ‘H=H’. But one cannot know that ‘H=P’ expresses a truth, without first knowing that the two reference-fixing descriptions in the rules (HES) and (PHOS) are uniquely satisfied by the same object, and such knowledge of course requires *a posteriori* empirical investigation.

As a consequence of our semantic hypothesis, it follows that various *logical* properties of the proposition that H=H (that is the proposition that H=P) can be hidden from view and not *a priori* accessible, even to one who has perfect understanding of the linguistic meanings of the relevant sentences. For instance the proposition that H=H *logically implies* the proposition that H=P (since every proposition logically implies itself), even though this logical fact cannot be known *a priori*. Also the proposition that H=P is a *logical truth*, since it is the same as the proposition that H=H (and the proposition that P=P), though again, this cannot be known through consideration of conceptual meanings alone. And finally, the proposition that H≠P is *not logically possible*, since it is identical with the logically impossible proposition that H≠H (and that P≠P), though again, one cannot know *a priori* that it’s not logically possible that H≠P.^{xvii}

This last case is especially important for the PQ-physicalist's defense of his view. For reasons that I will explain below, it certainly is *conceivable* that $H \neq P$. Surely everyone, whether Q-physicalist or type-dualist can agree with this. But as we've just seen, since the names 'Hesperus' and 'Phosphorus' are coreferential genuine terms, it turns out (*a posteriori*) that in fact it is *not* logically possible that $H \neq P$. Thus contrary to Chalmers' view, conceivability is no guarantee even of mere logical possibility. (See section 10 below for further discussion of this point.)

7. Why PQ-physicalism is false.

So the semantic facts about proper names and indexicals, based on the distinction between linguistic and propositional meaning, have the consequence that some of the most important logical properties of propositions can be hidden from *a priori* investigation, even for a person who has perfect understanding of the linguistic meanings of the sentences that express those propositions. That the propositions we express in ordinary language could have such hidden logical properties is of course one of the PQ-physicalist's crucial assumptions, since as we've seen, the PQ-physicalist claims that an instance of say '*X* feels a sharp pain' would in fact express a proposition that, unbeknownst to ordinary speakers, logically implies the existence of such things as brains and C-fibers.

But does the existence of support for this crucial assumption make it reasonable to believe even that PQ-physicalism *might* be true? Not really. The fact that some logical properties of some propositions can only be known *a posteriori* shows at most that there is perhaps room in logical space for the *sort* of semantic view that the PQ-physicalist wants to endorse. But what any PQ-physicalist owes the rest of us is an *actual account* of the semantics of phenomenal predicates, an account that is based on real semantic evidence and that actually explains how the PQ-

physicalist's incredible hypothesis - that phenomenal predicates actually ascribe complex, perhaps completely unknown, neurophysiological properties - could possibly be true.

Instead, PQ-physicalists can offer us only a mere *analogy* which they allege to hold between phenomenal predicates on the one hand, and indexical pronouns, proper names and (as we shall discuss below) natural kind terms, on the other. However, no defender of Q-physicalism has ever provided any evidence to support this analogy, and in fact the analogy breaks down in rather spectacular ways.

Imagine a tribe of prehistoric hominids who communicate in a rudimentary language about, among other things, each others' cognitive states and sensations. Their language contains a predicate which means exactly 'feels a sharp pain', but being quite ignorant of anatomy, these hominids have no concept of the brain and certainly no concept of C-fibers. Nevertheless, according to PQ-physicalism, when our hominids use their predicate to say that a given person feels a sharp pain, the proposition asserted is identical with the proposition that the C-fibers of the person's brain are firing to degree *n*. And so, even though our hominids *have no concepts* of either brains or C-fibers, they are capable of using their language to assert the proposition that the person has a brain whose C-fibers are firing to degree *n*. But surely, this consequence is just preposterous! Our hominids can obviously assert nothing about brains and C-fibers, *since they have no concepts of these things*.^{xviii}

Notice that there is nothing about the semantics of indexicals and proper names which at all supports the idea that the conceptual meanings of our words could be so radically distant from the logical implications of what we say as in the case of the hominids just described. One who says that H=H may indeed be unaware that she is saying something that logically implies that H=P. But there is nothing whose existence is logically implied by the proposition in question of

which an understanding speaker is incapable of conceiving. It is true in this case that an understanding speaker of ‘H=H’ who thereby asserts the proposition that H=P might be unable to conceive *that* H=P, since she might not have the concept expressed by ‘Phosphorus’.^{xix} Still, the speaker can conceive *of* Phosphorus, since she can conceive *of* Hesperus, and of course, Hesperus = Phosphorus. I am not suggesting that one who asserts a proposition with understanding must be able to conceive of that proposition in *every* possible way in which it *can* be conceived. Rather, I’m saying that if a person can assert with understanding a proposition that formally involves or immediately implies the existence of a certain kind of thing (such as brains or C-fibers), then the person must have at least *some* way of conceiving of that kind of thing.

To avoid a certain sort of misunderstanding, I should emphasize that I am *not* assuming in my argument that our primitive hominids have no concept of the property of having a brain whose C-fibers are firing to degree *n*. Since the hominids by assumption do have a concept of the property of feeling a sharp pain, and since the PQ-physicalist asserts that this property is identical with some neurophysiological property (such as that of having a brain whose C-fibers are firing to degree *n*), I would beg the question to just *assume* that our hominids have no concept of the latter property. But I am not making this assumption.

Rather, I am assuming only what should be common ground, namely, that our hominids have no concept that applies to all and only *brains*, and no concept that applies to all and only *C-fibers*. (I take it to be obvious that one could have the concept of a sharp pain, without having any concept whatever that picks out either brains or C-fibers.) Thus it is clear, and should be common ground, that our hominids cannot assert with understanding that anyone has a *brain*, or that anyone’s body contains *C-fibers*, since they have no concepts of either brains or C-fibers. Yet PQ-physicalism implies that our hominids *can* assert with understanding propositions to the effect

that the C-fibers of given persons' brains are firing to degree *n*. But again, this is just absurd. If one has no concepts of the relatively simple properties of being a brain or being a C-fiber, then one certainly cannot assert (with understanding) any proposition that predicates a complex property which in turn is constructed (say, truth-functionally) out of those simpler properties.^{xx}

What is absurd about the PQ-physicalist's semantic hypothesis is that according to this hypothesis, we have given meanings to some of the words of our language (namely, certain mental predicates) which allow us to use these words to in effect say things having logical implications that we cannot really understand or even conceive of. This is absurd because it violates the basic purpose for which we have language in the first place. We give meanings to our words so that we can communicate and say things which we understand, whose truth-conditions we can know in advance. Imagine the case of a person Mutt, who whimsically introduces a new predicate 'whatever' into his language, declaring that it is to be used to predicate whatever is his friend Jeff's favorite property. Not, you understand, that 'whatever' is to *mean* 'has Jeff's favorite property'; rather it is to *predicate* the relevant property, whatever it is. So suppose that the relevant property is that of being a bachelor. Then whenever Mutt utters a sentence of the form 'X is whatever', he's actually saying that a given object is a bachelor, and is thus asserting of that object that it is an unmarried male of marriageable age. But of course, neither Mutt nor anyone else (with the possible exception of Jeff) has the slightest idea that this is what Mutt is saying. But it is surely just absurd to suppose that we could use words with meanings such as this, since in doing so, we'd be using words with meanings that *prevent* ourselves and others from understanding what we're saying in using the words. But then we'd be undermining the very purpose for which we give meanings to words in the first place.

The PQ-physicalist's semantic hypothesis allows there to be an implausibly huge gap between one's understanding of an expression's linguistic meaning and one's understanding of what can be said, what sorts of propositions can be asserted, by use of that expression. Clearly there must be some *constraints* on the kind of semantic relations that can hold between linguistic and propositional meaning, constraints that the PQ-physicalist is wantonly flouting. We've seen that PQ-physicalism implies the absurd consequence that one who asserts the proposition that a given person feels a sharp pain would also be asserting a proposition that *immediately, formally*, implies the existence of kinds of things such as brains and C-fibers, of which the speaker might be completely unable to conceive. To block this sort of consequence, I would propose the following reasonable constraint:

(CONS)A semantic hypothesis must not imply that a speaker *S* could use a sentence with a given linguistic meaning, having perfect understanding of that meaning, to assert or express a proposition *p*, even though *p* formally implies the existence of objects of a given kind *K* such that *S* has no way to conceive of *K*s (where *p* is logically possible and it is not logically necessary that *K*s exist).^{xxi}

Here, by 'formally implies' I mean *strictly* logical implication, as opposed to the broadly logical notion that I explained earlier and that I've been calling 'logical implication'.^{xxii}

8. But what about H₂O?

Recent defenders of Q-physicalism seem to commonly assume that ordinary natural kind terms like 'water' predicate (or 'refer to') the same properties as complex theoretical terms like 'H₂O' which describe the physical natures of members of the relevant kinds. (See for instance Lycan, 1990, p. 120 and Loar, 1997, p. 599.) Of course, it goes without saying that the terms 'water' and 'H₂O' have quite different conceptual or linguistic meanings. So if these terms did

indeed predicate (or ‘refer to’) the same property, then PQ-physicalists would have an important additional analogy on which to base their claim that ordinary phenomenal predicates can ascribe complex physical properties to the brain.

However, it is easy to see that, for the same reason that the PQ-physicalist’s semantic hypothesis about phenomenal predicates is false, it is also false that ‘water’ and ‘H₂O’ predicate (or ‘refer to’) the same property. When one of our primitive hominids says, for instance, that a given gourd *g* contains water, she is not asserting the proposition that *g* contains H₂O. As before, she cannot assert this proposition with understanding since she has concepts of neither H₂O, hydrogen, nor oxygen. Hence the proposition that *g* contains water is not the same as the proposition that *g* contains H₂O: while the latter proposition formally implies the existence of H₂O molecules, hydrogen atoms, and oxygen atoms, the former proposition formally implies none of these things. Hence the property of being water is not the same as the property of being H₂O.

Q-physicalists who assume that ‘water’ and ‘H₂O’ predicate the same property may be laboring under the impression that this result was ‘shown’ by Kripke’s (1972) and Putnam’s (1975) seminal discussions of the semantics of natural kind terms. This is just wrong. All the semantic evidence adduced by Kripke and Putnam can be accounted for by a semantic theory that fails to have this alleged consequence. For instance, I have proposed elsewhere (McKinsey 1987) that the linguistic meaning of ‘water’ is provided by a ‘property-fixing’ rule of the following sort:

(W) For any token Φ of ‘is water’, and any property F, Φ is to predicate F if and only if there is a unique natural kind K such that (in the actual world) the watery stuff found in our environment belongs to K, and F = the property of belonging to K.

Here, ‘watery stuff’ is a euphemism for a conjunction of surface qualities that ordinary speakers associate with ‘water’. Use of the indexical expression ‘our environment’ allows me to

distinguish the meaning that ‘water’ has in the English spoken by *us*, the inhabitants of Earth, from the meaning of ‘water’ in the English spoken by our counterparts on the Twin Earth of Putnam’s (1975) famous example.^{xxiii}

According to (**W**), the *propositional* meaning of ‘water’ is the property of belonging to K, where K is the natural kind to which samples of water actually belong. But notice that the property of belonging to K is *not* the same as the property of being composed of H₂O molecules, for the simple reason that having the former property does not logically imply having the latter. There is to be sure an important relation between these two properties, but it is not that of *identity*. Rather, it is that of (metaphysically) necessary equivalence. What scientists discovered is that being water, that is, belonging to K, is necessarily determined by being composed of H₂O molecules. Since membership in a given natural kind is an essential property of anything that has it, it has turned out to be a (metaphysically) necessary truth that water is composed of H₂O molecules.

Another reason why some Q-physicalists may have assumed that ‘water’ and ‘H₂O’ predicate (or ‘refer to’) the same property is that they have assumed that the necessarily true sentence ‘Water is H₂O’ is an *identity*-sentence. Of course, if this were a true identity-sentence, then it would have to express the necessary identity of the type *water* with the type *H₂O*, and so the truth of the sentence would imply that ‘water’ and ‘H₂O’ refer to the same types (properties). But again, this is simply wrong. The sentence ‘Water is H₂O’ does not (falsely) assert that the types *water* and *H₂O* are identical. For this sentence does not involve reference to, is not *about*, abstract types at all. Rather, the sentence is about *water*, the liquid *stuff* that we drink and bathe in, and it says simply that all stuff that is water (i.e., that belongs to K) is composed of H₂O molecules. (See McKinsey 1987, p. 31 and Burge 1972.)^{xxiv}

So we may conclude that, like the semantics of indexicals and proper names, the semantics of natural kind terms also provides no support for the PQ-physicalists' semantic hypothesis about phenomenal predicates.

9. Why RQ-physicalism is false.

I have argued that PQ-physicalism is false on the grounds that when we use mental predicates to ascribe sensory qualities, the propositions we express and assert do not formally imply the kinds of physical consequences which according to PQ-physicalism these propositions must imply. Now PQ-physicalism has this kind of false consequence only because it holds that the mental predicates in question *ascribe* or *predicate* sensory qualities that are identical with complex physical properties of the brain. But as I mentioned earlier, it is not altogether clear that Q-physicalists like Loar and Perry would agree with PQ-physicalism, since they tend to use the vocabulary of *reference*, rather than predication, to state their view. ^{xxv}

So we should consider RQ-physicalism, a view which does not claim that phenomenal predicates ascribe or predicate properties that turn out to be physical properties of the brain. Rather, RQ-physicalism claims that such mental predicates somehow involve *direct reference* to sensory qualities of experience, where these qualities turn out to be identical with complex physical properties of the brain. RQ-physicalism is a much more abstract and inexplicit sort of view than PQ-physicalism, which is quite explicit and straightforward about what it's claiming. PQ-physicalism just takes phenomenal predicates and concepts at face value, assumes that these predicates and concepts predicate or ascribe sensory qualities, and then adds that the qualities predicated happen to be complex physical properties of the brain.

By contrast, RQ-physicalism says that our uses of phenomenal predicates and concepts somehow involve direct reference to sensory qualities that turn out to be complex physical

properties of the brain, but it does not explain how this direct reference (as opposed to predication) goes, and it is quite inexplicit as to how any mental predicates could have such a logical form that there is only direct reference to, rather than predication of, the relevant sensory/physical properties. But abstract and inexplicit as it is, RQ-physicalism has interesting logical features that make it worth discussing. In particular, RQ-physicalism has the distinct advantage of failing to imply any of the false consequences that I used to argue against PQ-physicalism.

For definiteness, let's consider the possibility that a mental attribution of the form

(P) X feels a sharp pain

might express a proposition of the form ' X has N ', where N is a singular term, such as a proper name or demonstrative pronoun, which refers directly to a certain sensory quality (namely, the one allegedly mentioned in the predicate contained in (P)). In fact, let's just give this quality a name, say, 'Charley'. Then the idea is that instances of (P) express propositions of the form

(C) X has Charley.

Of course, the RQ-physicalist adds that the name 'Charley' refers to a sensory quality that in fact turns out to be a complex physical property, perhaps the very same neurophysiological property as the one that is *predicated* by instances of our old friend

(B) The C-fibers of X 's brain are firing to degree n .

But there is a big difference between propositions of the form (C) and propositions of the form (B). For unlike propositions of the form (B), propositions of the form (C) do not formally imply the existence of such things as brains and C-fibers. This is not implied, even if the name 'Charley' contained in (C) refers to what is in fact a complex physical property. For instances of (C) say only that a given object X bears the relation of *having* (of instantiating or exemplifying) to

that property. Since unlike instances of (B), instances of (C) do not ascribe or predicate the physical property in question, instances of (C) do not say or formally imply anything about what that property is, or about what its physical nature is. So instances of (C) do not formally imply the existence of brains, or C-fibers, or any other sort of physical object.

Thus a form of RQ-physicalism which says that mental attributions like (P) express propositions of the form (C) that are *about* sensory/physical properties appears to be a view that is perfectly consistent with all Cartesian intuitions to the effect that mental attributions in general fail to logically imply any physical propositions. Thus my earlier objections to PQ-physicalism fail to apply to RQ-physicalism. Still, I think it is fairly easy to see that RQ-physicalism is false, for the simple reason that it gives a false account of the semantics of most phenomenal predicates, such as the predicate contained in (P), ‘feels a sharp pain’. This predicate, just like other predicates, neither refers to, nor contains any other terms that refer to, properties of any sort. Rather, predicates like this, just like other predicates, are used to predicate and ascribe properties, not to refer to them.

One important difference between predicates and referring expressions like proper names and demonstratives, is that unlike referring expressions, predicates can be *modified* by adjectives and adverbs, so that as a result, predicates can have complex structures and bear important logical relations to each other. RQ-physicalism cannot do justice to this simple fact. Consider for instance the RQ-physicalist’s suggestion that instances of (P) (‘*X* feels a sharp pain’) have the same logical form as instances of (C) (‘*X* has Charley’). If this suggestion were true, then it would be impossible to explain why instances of (P) all logically imply instances of

(Q) *X* feels a pain.

For if we apply the RQ-physicalist's hypothesis about (P) to the case of (Q), then instances of (Q) would have the logical form of

(D) *X* has Dudley,

where 'Dudley' names some property distinct from the property named by 'Charley'. But of course no instance of (C) can logically imply an instance of (D), given that the imbedded names have distinct properties as referents. Even if we assume that 'Charley' names the property of feeling a sharp pain and 'Dudley' names the property of feeling a pain, the implication cannot go through, since (C) and (D) function to effectively *hide* the logical structures of any properties that they might name.^{xxvi}

Recall that for the RQ-physicalist, the main advantage of the suggestion that phenomenal predicates might refer to properties in the manner of names or demonstratives, is that this suggestion blocks the formal implication of any implausible physical consequences that uses of phenomenal predicates clearly do not have. But the great *disadvantage* of this same suggestion is that it not only blocks bad, unwanted, logical implications, but it *also* blocks the logical implications that uses of phenomenal mental predicates really do have.

So Q-physicalists face a serious dilemma. They can endorse PQ-physicalism, accepting the fact that phenomenal concepts and predicates can be used to predicate and ascribe complex mental properties, thus doing justice to the fact that phenomenal mental predicates can bear important logical relations to each other. On this alternative, the Q-physicalist must identify the property ascribed by a predicate like 'feels a sharp pain' with a complex physical property of the brain. But then it follows that ordinary ascriptions of sensory qualities to persons and experiences must express complex physical propositions about the brain, so that when we make such

ascriptions we are asserting propositions that formally imply the existence of kinds of things of which we may be incapable of conceiving. This consequence, I take it, defies credulity.

On the other hand, the Q-physicalist can avoid this absurd consequence by instead endorsing RQ-physicalism, insisting that phenomenal mental predicates don't predicate complex physical properties, but rather merely *refer* to such properties. But then, as we've just seen, the Q-physicalist's view conflicts with the fact that phenomenal mental predicates can and do in fact ascribe complex mental properties that bear important logical relations to each other.

Whichever of these two sorts of view the Q-physicalist chooses to endorse, the view endorsed is false. Hence, Q-physicalism is false.

I have never seen these two forms of Q-physicalism distinguished from each other, so it may well be that the defenders of Q-physicalism have simply *conflated* the two views, thus tacitly assuming that they could have the advantages of both views and the disadvantages of neither. A conflation of this kind is of course fostered by the Q-physicalists' unfortunate tendency to say that phenomenal concepts and predicates *refer to* (rather than predicate) sensory qualities. So the mistake at the root of Q-physicalism may simply be a confusion of *reference* and *predication*.

10. Chalmers' concept of logical possibility.

In the above discussion, chiefly in my argument against PQ-physicalism, I have relied crucially on the notion of logical possibility that I explained earlier. In this respect my treatment of these issues resembles that of Chalmers (1996), who also relies on a concept of logical possibility in his arguments against reductive physicalism. But my concept of logical possibility is significantly different from Chalmers'. One major difference is that for me, logical possibility is a property of *propositions*, whereas for Chalmers, it is a property of what he calls 'statements'. Although he does not say so explicitly, Chalmers seems to take statements to be a species of

sentence, since as he uses the term, ‘statements’ are said to have meanings and to have words as parts.

Chalmers’ concept of a sentence’s being logically possible is based on a distinction that he makes between two types of meaning. In his terminology, every word (or ‘concept’) and every sentence (or ‘statement’) has both a *primary* intension and a *secondary* intension. This distinction is the same as, or at least very similar to, the distinction that I described earlier between linguistic meaning (primary intension) and propositional meaning (secondary intension). In effect, from my point of view, a sentence (or ‘statement’) is said by Chalmers to be logically possible if and only if it satisfies one of two conditions: either (i) the proposition that the sentence actually expresses (the sentence’s secondary intension) is true in some possible world, or (ii) the sentence’s linguistic meaning (its primary intension) is such that, in some possible world w , if the sentence were used in w with that meaning, it would express a proposition that is true in w . For instance, the sentence ‘Water is not H₂O’ actually expresses a necessary falsehood, and so it does not satisfy the first of the above two conditions (its ‘secondary intension’ is not true in any possible world). Nevertheless, Chalmers counts ‘Water is not H₂O’ as logically possible, because it could be used with its actual linguistic meaning (its ‘primary intension’) to express a truth. For instance, in a possible world in which the watery stuff in our environment is not composed of H₂O molecules but rather of, say, XYZ molecules, ‘Water is not H₂O’ would express a truth, and so Chalmers counts the sentence as logically possible.

I would also count the sentence ‘Water is not H₂O’ as logically possible, but for a quite different reason. Extending my notion from propositions to sentences, I would say that the sentence is logically possible because the proposition that the sentence *actually* expresses (its

‘secondary intension’) is logically possible. This proposition in turn is logically possible because it is logically consistent, that is, does not logically imply a contradiction.

Both Chalmers’ concept of logical possibility and mine count ‘Water is not H₂O’ as logically possible, though for different reasons. But in other cases, the two concepts diverge in their application. As we saw earlier, it is logically possible in my sense that Descartes is not a physical object. But Chalmers provides no sense in which the sentence ‘Descartes is not a physical object’ is logically possible. For on the one hand, the proposition actually expressed is true in no possible world (since Descartes is in fact a physical object, as Chalmers and I agree). On the other hand, there also seems to be no world in which the sentence *could* express a truth, given its actual linguistic meaning. The reason why ‘Water is not H₂O’ can express a truth is that the sort of contingent description that fixes the reference of ‘water’ (e.g., ‘the watery stuff in our environment’) can, in some non-actual but possible world, refer to stuff that is not H₂O. But as I’ve argued at length elsewhere, most ordinary names like ‘Descartes’ (and unlike ‘Hesperus’ and ‘Phosphorus’) have no descriptive meanings in the public language. (See McKinsey 1984, 1994, 1999.) So there is in this case no distinction between the proposition actually expressed by ‘Descartes is not a physical object’ and other propositions that the sentence could express, given its actual meaning.^{xxvii} For we have to either take this sentence to have *no* linguistic meaning (no primary intension), since it contains a name that has no meaning in the language, or we have to take the sentence’s linguistic meaning to be a function of the actual *referent* of ‘Descartes’, in which case the sentence’s primary and secondary intensions are identical. In either case, the sentence expresses a truth in *no* possible world given its actual meaning, and so the sentence is not logically possible in Chalmers’ sense.

Now I take it that a sentence like ‘Descartes is a physical object’ expresses a paradigm of *a posteriori* metaphysical necessity. Hence its negation should be taken to express a paradigm of something that is metaphysically but *not* logically impossible. Since my concept of logical possibility implies, but Chalmers’ does not, that it is logically possible that Descartes is not a physical object, it seems to me that my concept has a distinct advantage over his.

We’ve just seen that something can be logically possible in my sense without being logically possible in Chalmers’ sense. It also happens that something can be logically possible in Chalmers’ sense without being logically possible in mine. As we saw earlier, since the proposition that $H \neq P$ is the same as the proposition that $H \neq H$, it follows on my account that it is *not* logically possible that $H \neq P$. But on Chalmers’ account, this is logically possible. For given that the linguistic meanings (primary intensions) of the (unusual descriptive) names ‘Hesperus’ and ‘Phosphorus’ are provided by reference-fixing semantic rules like **(HES)** and **(PHOS)** above, the sentence ‘ $H \neq P$ ’ will express a truth in any possible world in which the relevant contingent reference-fixing descriptions are satisfied by distinct planets. Chalmers agrees with me that the names in question may well have their references fixed by description in this way. So it follows from his account that it is logically possible that $H \neq P$.

The ‘two-dimensional’ semantics appealed to by Chalmers in his account of logical possibility is similar to, if not the same as, the semantic view of singular and natural kind terms that I described earlier and that I’ve been defending for many years. (See McKinsey 1978a, 1978b, 1984, 1987, 1994, and 1999.)^{xxviii} So it’s interesting that Chalmers should use this two-dimensional semantics to motivate a concept of logical possibility that is so different from the concept that I’ve proposed above and elsewhere (McKinsey 1991a and 1991b). Again, on my account, a sentence expresses a logical possibility just in case the proposition that the sentence

actually expresses is logically consistent. Whether the sentence in question, like ‘H≠P’, *could* express a truth given its actual meaning is for me strictly irrelevant to whether the sentence does or does not in fact express a logical possibility. To be sure, that such a sentence *could* express a truth may explain why that sentence *appears* to express a logical possibility, but in my view, it’s a mistake to infer, as Chalmers seems to do, that the sentence actually does express a logical possibility in any interesting sense.

On the other hand, if a sentence *p* could express a truth given its actual meaning, then in my view, it does indeed follow that it is *conceivable* that *p*. On the theory of cognitive ascriptions that I’ve defended elsewhere (McKinsey 1986, 1994, and 1999), the *linguistic* meaning of a given sentence *p* can semantically determine what cognitive property is expressed by cognitive predicates that contain *p* (such as ‘thinks that *p*’ and ‘believes that *p*’), and this can be determined *independently* of the proposition that the sentence *p* actually expresses. In particular, on my theory, this happens when the relevant sentences contain (unusual) proper names that have descriptive meanings, meanings of the kind provided for the names ‘Hesperus’ and ‘Phosphorus’ by such reference-fixing rules as (HES) and (PHOS).

Thus, in the actual world, if one conceives (thinks) that H≠P, then the content of one’s thought is in fact an inconsistent proposition. But in another possible world *w*, one could have *exactly the same thought*, also thinking in *w* that H≠P, and be thinking something true, because the content of one’s thought in *w* is a different, consistent proposition that is true in *w*. The proposition in question would of course be the proposition expressed in *w* by ‘H≠P’, given the sentence’s actual (descriptive) meaning. Thus, on my theory of cognitive ascriptions, it is conceivable that H≠P, because it is possible to truly conceive (think) that H≠P. (See McKinsey 1999, especially pp. 351-2, note 13.)

In general, it is conceivable that p just in case it is possible to consistently conceive (think) that p , where there are two different ways in which one can consistently conceive (think) that p : either (i) the proposition that p (the proposition actually expressed by ' p ') is in fact logically possible in my sense, or (ii) there is a possible world w in which one could have the very same thought that p , where one's thought that p is true in w . Note the structural similarity of this account of conceivability to Chalmers' account of logical possibility. Also note that on this account, in contrast to that of Chalmers, its being conceivable that p is necessary but *not* sufficient for its being logically possible that p .

11. Chalmers' zombies.

The argument against reductive physicalism on which Chalmers seems chiefly to rely is based on the logical possibility of *zombies*. A 'zombie' is by definition a creature that is physically, molecule-for-molecule, identical with some actual human being, but that completely lacks *consciousness* and so never has any mental experiences characterized by sensory qualia (there is nothing that it is like to be a zombie). Now by my lights, it is indeed logically possible that there are zombies, at least in my sense of 'logically possible'. For it seems to me that even the totality of physical facts that I for instance would share in common with my zombie twin, would be insufficient to logically imply any mental facts about my conscious states. It follows of course that none of these mental facts about me are identical with any physical facts, since if any mental fact about me *were* identical with a physical fact, then it *would* be logically implied by the totality of physical facts about me (for the simple reason that this mental fact would be one of the totality of physical facts and would logically imply itself).

But while I agree that this simple argument is sound, I'm afraid that it's not very persuasive. A defender of reductive physicalism like the PQ-physicalist would no doubt find the argument question-begging, since the PQ-physicalist's view is that mental facts about a person's conscious states just *are* physical facts about the brain. This is one reason why I prefer my earlier arguments against Q-physicalism to the zombie argument. Recall that my earlier argument against PQ-physicalism does not merely *claim* that facts about conscious states fail to logically imply any physical facts about the brain. Rather, my argument provides a reason for this claim, by pointing out that if facts about conscious states logically imply physical facts because the former facts are *identical* with complex physical facts about the brain, then pedestrian facts about conscious states that people state every day would (absurdly) formally imply the existence of kinds of things of which we may not even be able to conceive.

Notice that my argument and the zombie-argument in effect proceed in opposite directions. The zombie-argument contends that the physical facts can (logically) occur in the absence of facts about consciousness. My argument contends conversely that facts about consciousness can (logically) occur in the absence of the physical facts. At one point (1996, pp. 147-8) Chalmers contends that this difference makes the zombie-argument preferable to arguments like mine. It is true, as he points out, that the zombie-argument is more *general* in its application. While arguments like mine, if sound, refute only *identity* theories like PQ-physicalism, the zombie-argument, if sound, refutes both identity theories and weaker forms of reductive physicalism on which the mental facts merely logically *supervene* upon the physical facts, such a theory as

(LS) The totality of physical facts which hold in the actual world *logically implies* all of the mental facts which also hold in the actual world.

(See Chalmers 1996, pp. 41-2.)

But the fact that the zombie-argument, if sound, refutes both identity theories like PQ-physicalism and supervenience theories like (LS) does not, it seems to me, give the zombie-argument any distinct advantage over arguments like mine. For one thing, as I say, the zombie-argument is not very persuasive. For another, the only form of reductive physicalism yet proposed that is not an identity theory and that is strong enough to entail (LS) is *functionalism*, and as I remarked earlier, there are ample independent grounds for rejecting functionalism.

I should emphasize that so far, I've stated the zombie-argument using only *my* sense of logical possibility. The argument is weak when understood this way, but it is *much weaker* if understood in terms of Chalmers' concept of logical possibility. In order for it to be logically possible in Chalmers' sense that there are zombies, it must be true *in some possible world* that there are zombies, given either the primary or secondary intensions of 'there are zombies'. I admit to having some difficulty distinguishing the primary and secondary intensions of this sentence, and so I strongly suspect that the intensions are identical in this instance. But in any case, it seems to me that it is very implausible, or at least highly controversial, to suppose that there is a possible world in which it is in any sense true that there are zombies.

Consider my zombie twin and me at the moment when each of us stubs his toe on a chair leg. Our toes are damaged in exactly the same way, and as a result, exactly the same nerve impulses are transmitted to our brains, which end up being in exactly the same state. Only of course, while we are in molecule-for-molecule identical brain states, I feel excruciating pain in my toe while my zombie twin feels nothing! This *seems* to be at least *logically* possible in my sense: we certainly cannot *deduce*, using logic alone, that my zombie twin must feel pain and

hence that the scenario is contradictory. But is it *really possible* that he feels no pain? In describing this scenario have we *really* described part of a possible world? I very much doubt it.

Chalmers, I should emphasize, never gives any good reason at all to believe that there are possible zombie worlds. He only contends (correctly) that it is *conceivable* that there are zombies, pointing out frequently (and correctly) that there is no *incoherence* or *inconsistency* in the supposition that there are zombies (1996, pp. 94-99). But this of course only supports the conclusion that zombies are logically possible in *my* sense. It does not at all support the conclusion that in some possible world it is true that there are zombies. One might as well argue, repeating Descartes' error, that since it is not contradictory to suppose that Descartes is not a physical object, there is *therefore* some possible world in which Descartes exists as a purely spiritual being!^{xxix}

The problem is that Chalmers' sense of logical possibility forces him to confuse logical and metaphysical possibility, whether he wants to or not. In his sense, the contention that zombies are logically possible entails that there are possible zombie worlds, and since the possible worlds = the metaphysically possible worlds, it follows that zombies are metaphysically possible. But this is just too much to swallow.^{xxx} For the claim that zombies are metaphysically possible contradicts not only all the forms of physicalism that we've so far considered, it also contradicts a view that some property dualists including myself believe may well be the truth about the relation between the mental and the physical. This is the view that while mental properties, propositions, and facts are neither identical with nor otherwise logically reducible to physical properties, propositions, and facts, the mental may nevertheless *metaphysically supervene* upon the physical. A simple way to express this view would be:

(MS) The totality of physical facts which hold in the actual world *metaphysically entails* all of the mental facts which also hold in the actual world.

If we count (MS) as a form of physicalism, then it is certainly the weakest form we've yet considered. That is, (MS) is entailed by all of the other forms, although it entails none of them. I am myself strongly inclined to believe that (MS) is true. Certainly, (MS) deserves further serious consideration of the sort recently provided by Nagel (2000).

Chalmers' zombie-argument is unconvincing because it proves too much. The assumptions required by the argument are so strong and controversial that they contradict not only clearly *false* forms of reductive physicalism like Q-physicalism, but also non-reductive forms of physicalism like (MS) that are likely to be true. By contrast, when the zombie-argument is understood in terms of my sense of logical possibility, its premises are perfectly consistent with non-reductive views like (MS).

12. Conclusion.

Q-physicalists invoke the distinction between concepts and properties, a special case of the distinction between linguistic and propositional meaning, as a way of countering plausible Cartesian intuitions to the effect that propositions which ascribe sensory qualities to our experiences do not logically imply the existence of such complex physical things as brains, neurons, and C-fibers. PQ-physicalists in particular maintain that propositions ascribing sensory qualities must logically imply complex physical propositions about the brain, where these implications are hidden from any *a priori* investigation that is based on understanding of conceptual meanings alone.

I have agreed that there is a grain of truth underlying the PQ-physicalist's position: research in referential semantics concerning indexical pronouns, proper names, and natural kind terms does show that some of the logical properties of the propositions we assert by use of such terms can be hidden from *a priori* investigation. However, I've argued that the PQ-physicalist goes too far. For if PQ-physicalism were true, then many commonplace ascriptions of sensory qualities to experiences would formally imply the existence of kinds of things of which the speakers cannot, and perhaps never could, conceive. As a remedy for the PQ-physicalists' semantic excesses, I proposed (CONS), a reasonable constraint on how linguistic meaning can be related to propositional meaning.

A second version of Q-physicalism, RQ-physicalism, claims that we never use mental concepts and predicates to *ascribe* or *predicate* sensory qualia; rather, we only use such concepts and predicates to *refer* to sensory qualia, where these qualia turn out in fact to be complex physical properties of the brain. RQ-physicalism does avoid the semantic excesses and absurd consequences of PQ-physicalism, but it does so only at the expense of implying the false consequence that phenomenal mental predicates are never used to ascribe complex mental properties that can bear important logical relations to each other.

Since both forms of Q-physicalism are false, Q-physicalism is false. So the phenomenal qualities of sense experience are not identical with any physical properties of the brain. Nor, I take it, do such qualities logically supervene upon the physical properties of the brain, or indeed upon any other physical properties. Hence it seems to me that no form of *reductive* physicalism is true. Phenomenal properties, as well as the propositions and facts that involve such properties, exist 'over and above' the physical properties, propositions and facts. But again, it does nevertheless seem likely that such non-physical phenomenal properties, propositions, and facts are somehow

metaphysically necessitated or entailed by physical properties, propositions, and facts about the brain. But I confess that I do not have the slightest idea of how or why this could be so.^{xxxii}

REFERENCES

- Anchustegui, Ann-Marie: 1997, *Qualitative Content and the Mind-Body Problem* (Ph.D. dissertation, Wayne State University).
- Block, Ned: 1978, "Troubles with Functionalism," in C.W. Savage (ed.) *Minnesota Studies in the Philosophy of Science*, Vol. 9 (Minneapolis: University of Minnesota Press).
- Burge, Tyler: 1972, "Truth and Mass Terms," *Journal of Philosophy* 69, 263-282.
- Burge, Tyler: 1988, "Individualism and Self-Knowledge," *Journal of Philosophy* 85, 649-663.
- Cartwright, Richard: 1961, "Propositions," in R. J. Butler (ed.) *Analytic Philosophy*, Vol. 1 (Oxford: Basil Blackwell).
- Chalmers, David J.: 1996, *The Conscious Mind* (Oxford: Oxford University Press).
- Chalmers, David J.: 1999, "Materialism and the Metaphysics of Modality," *Philosophy and Phenomenological Research* 59, 473-96.
- Chalmers, David J.: 2002a, "The Components of Content," in *Philosophy of Mind: Classical and Contemporary Readings*, edited by D. J. Chalmers (Oxford: Oxford University Press), 608-633.
- Chalmers, David J.: 2002b, "Does Conceivability Entail Possibility?" in T. Gendler and J. Hawthorne (eds.) *Conceivability and Possibility* (Oxford: Oxford University Press), 145-200.
- Conee, Earl: 1994, "Phenomenal Knowledge," *Australasian Journal of Philosophy* 72, 136-150.
- Davies, Martin and Humberstone, Lloyd: 1980, "Two Notions of Necessity," *Philosophical Studies* 38, 1-30.
- Descartes, René: 1977, *The Philosophical Works of Descartes*, Volumes I and II, translated by E. Haldane and G. Ross (Cambridge, England: Cambridge University Press).
- Frege, Gottlob: 1892, "On Sense and Reference," in *Translations from the Writings of Gottlob Frege*, P. Geach and M. Black (eds.), (Oxford: The Oxford University Press, 1966).
- Frege, Gottlob: 1956, "The Thought: A Logical Inquiry," translated by A.M. Quinon and A. Quinon, *Mind* 65, 289-311.
- Goldman, Alvin I.: 1993, "The Psychology of Folk Psychology," *Behavioral and Brain Sciences* 16, 15-28.
- Hill, Christopher S.: 1997, "Imaginability, Conceivability, Possibility and the Mind-Body Problem," *Philosophical Studies* 87, 61-85.
- Hill, Christopher S., and McLaughlin, Brian P.: 1999, "There Are Fewer Things in Reality Than Are Dreamt of in Chalmers's Philosophy," *Philosophy and Phenomenological Research* 59, 445-54.
- Hooker, Michael: 1978, "Descartes' Denial of Mind-Body Identity," in M. Hooker (ed.), *Descartes: Critical and Interpretive Essays* (Baltimore: Johns Hopkins Press), 171-185.
- Jackson, Frank: 1982, "Epiphenomenal Qualia," *Philosophical Quarterly* 32, 127-136.
- Jackson, Frank: 1986, "What Mary Didn't Know," *Journal of Philosophy* 83, 291-295.

- Kaplan, David: 1977, "Demonstratives," in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan* (Oxford: Oxford University Press, 1989), 481-563.
- Kaplan, David: 1978, "Dthat," in P. Cole (ed.), *Syntax and Semantics*, Vol. 9, *Pragmatics* (New York: Academic Press), 221-253.
- Kaplan, David: 1979, "On the Logic of Demonstratives," in P. French, T. Uehling, and H. Wettstein (eds.), *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: University of Minnesota Press), 401-412.
- Kripke, Saul: 1972, "Naming and Necessity," in D. Davidson and G. Harman (eds.), *Semantics of Natural Language* (Dordrecht: D. Reidel), 253-355.
- Levine, Joseph: 1993, "On Leaving Out What It's Like," in M. Davies and G. Humphreys (eds.), *Consciousness: Psychological and Philosophical Essays* (Oxford: Blackwell's).
- Loar, Brian: 1990, "Phenomenal States," *Philosophical Perspectives* 4, Action Theory and Philosophy of Mind, 81-108.
- Loar, Brian: 1997, "Phenomenal States," in *The Nature of Consciousness*, N. Block, O. Flanagan, and G. Güzeldere (eds.) (Cambridge, Mass: The MIT Press), 597-616. [A revised version of Loar 1990.]
- Lycan, William G.: 1990, "What is the 'Subjectivity' of the Mental?" *Philosophical Perspectives* 4, Action Theory and Philosophy of Mind, 109-130.
- McKinsey, Michael: 1978a, "Names and Intentionality," *The Philosophical Review* 87, 171-200.
- McKinsey, Michael: 1978b, "Kripke's Objections to Description Theories of Names," *Canadian Journal of Philosophy* 8, 485-497.
- McKinsey, Michael: 1984, "Causality and the Paradox of Names," *Midwest Studies in Philosophy* 9, 491-415.
- McKinsey, Michael: 1986, "Mental Anaphora," *Synthese* 66, 159-175.
- McKinsey, Michael: 1987, "Apriorism in the Philosophy of Language," *Philosophical Studies* 52, 1-32.
- McKinsey, Michael: 1991a, "Anti-Individualism and Privileged Access," *Analysis* 51, 9-16.
- McKinsey, Michael: 1991b, "The Internal Basis of Meaning," *Pacific Philosophical Quarterly* 72, 143-169
- McKinsey, Michael: 1994, "Individuating Beliefs," *Philosophical Perspectives* 8, Logic and Language, 303-330.
- McKinsey, Michael: 1999, "The Semantics of Belief Ascriptions," *Noûs* 33, 519-557.
- Nagel, Thomas: 1974, "What Is It Like to be a Bat?" *Philosophical Review* 82, 435-450.
- Nagel, Thomas: 2000, "The Psychophysical Nexus," in P. Boghossian and C. Peacocke (eds.), *New Essays on the A Priori* (Oxford: Clarendon Press), 431-471.
- Perry, John: 1977, "Frege on Demonstratives," *The Philosophical Review* 86, 474-497.
- Perry, John: 1979, "The Problem of the Essential Indexical," *Noûs* 13, 3-21.
- Perry, John: 1997, "Indexicals and Demonstratives," in R. Hale and C. Wright, eds., *Companion to the Philosophy of Language* (Oxford: Blackwell's), 586-612.
- Perry, John: 2001, *Knowledge, Possibility, and Consciousness* (Cambridge, Mass.: The MIT Press).
- Plantinga, Alvin: 1970, "World and Essence," *The Philosophical Review* 79, 461-492.
- Plantinga, Alvin: 1974, *The Nature of Necessity* (Oxford: Oxford University Press).
- Pollock, John: 1982, *Language and Thought* (Princeton: Princeton University Press).

- Putnam, Hilary: 1975, "The Meaning of 'Meaning'," in K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science* 7 (Minneapolis: The University of Minnesota Press), 131-193.
- Russell, Bertrand: 1912, *The Problems of Philosophy* (New York: Oxford University Press, 1959).
- Seegerberg, Krister: 1973, "Two-dimensional Modal Logic," *Journal of Philosophical Logic* 2, 77-96.
- Soames, Scott: 2002, *Beyond Rigidity* (Oxford: Oxford University Press).
- Stalnaker, Robert C.: 1978, "Assertion," in P. Cole (ed.), *Syntax and Semantics*, Vol. 9, *Pragmatics* (New York: Academic Press), 315-322.
- Strawson, P.F.: 1950, "On Referring," *Mind* 59, 320-344.
- Tichý, Pavel: 1983, "Kripke on Necessity A Posteriori," *Philosophical Studies* 43, 225-241.
- Yablo, Stephen: 1993, "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research* 53, 1-42.
- Yablo, Stephen: 1999, "Concepts and Consciousness," *Philosophy and Phenomenological Research* 59, 455-63.

NOTES

-
- i. Others endorsing or at least suggesting this sort of strategy include Lycan 1990, Levine 1993, Conee 1994, and Hill 1997.
- ii. While holding a type-identity theory for sensory qualia, the Q-physicalist need not hold that *all* mental properties are identical with physical properties. For instance, some Q-physicalists might want to endorse a mixed form of reductive physicalism, holding that while sensory qualia are identical with physical properties of the brain, *cognitive* mental properties are multiply realizable functional properties that logically supervene upon, but are not identical with, physical properties.
- iii. See for instance the classic refutations of functionalism by Block 1978 and Goldman 1993.
- iv. The argument I have in mind occurs in *Meditation VI*. See Descartes 1977, Volume I, p. 190. I discussed this argument in McKinsey 1991a and 1991b. See also Burge 1988. My version of Descartes' argument was inspired by an argument that Alvin Plantinga once proposed, defended, and attributed to Descartes. (See Plantinga 1970, pp. 485-86 and 1974, pp. 66-69.) A similar version of Descartes' argument, also inspired by Plantinga, was proposed and usefully discussed by Michael Hooker (1978). (Thanks to Bill Stine for referring me to Hooker's paper.)
- v. This is the same as the definition that I gave of 'conceptually implies' in McKinsey, 1991a, p. 14, and 1991b, p. 152.
- vi. For suppose that it is not logically possible that $(p \ \& \ \sim q)$. Then $(p \ \& \ \sim q)$ logically implies a contradiction, and so it is a logical truth that if p then q ; but then p logically implies q (since in conjunction with the logical truth that if p then q , p immediately implies q by *modus ponens*). So

if p does *not* logically imply q , then it *is* logically possible that $(p \ \& \ \sim q)$. I make implicit use of this principle at various points in the text.

vii. For Arnauld's objection, see Descartes 1977, Volume II, *Objections IV*, p. 81.

viii. Thus my distinction is similar to Chalmers', who applies his distinction only to what he calls 'statements' and not to worlds (1996, p. 68). But apparently, Chalmers' 'statements' are just *sentences*, or perhaps sentences as used assertively, so that his distinction unlike mine applies to neither propositions nor worlds. I'll discuss the differences between Chalmers' distinction and mine more fully in sections 10 and 11 below.

ix. See for instance Chalmers 1996 and Hill and McLaughlin 1999.

x. This leads me to strongly suspect that the notion of 'logical possibility' as I'm understanding it, the notion that Descartes discovered, even though it is a perfectly fine and useful notion, may not correspond to any actual meaning of the word 'possible'. Perhaps, then, it would be less confusing or misleading if we used other terminology instead, such as 'logically consistent'. However, since the expression 'logically possible' in the sense I'm using it has become so well entrenched, I'll continue to use it this way, while recommending caution. In particular (unfortunately), it should not be assumed that 'it is logically possible that p ' entails 'it is possible that p '.

xi. To avoid trivial satisfaction by contradictory properties, we should add that the property is one that it is logically possible for an object to possess.

xii. Strictly, these definitions are more naturally taken to define what we might call *basic* physical propositions and facts, with the non-basic physical propositions and facts being either truth-functions or generalizations of the basic ones. Ignoring this distinction should not affect the discussion to follow.

xiii. See also Pollock, 1982. The basic distinction at the level of whole sentences had earlier been made in the seminal papers by Strawson (1950) and Cartwright (1961). The most thorough and plausible recent treatment of indexicality that I know of is Perry 1997. My distinction between linguistic and propositional meaning is very similar to, though slightly more general in its application than, Kaplan's (1977) famous distinction between *character* and *content*. My distinction is also (as far as I can tell) the same as Perry's distinction between *reflexive* and *subject-matter* content. See Perry 2001, Chapter 3.

xiv. This consequence stands in sharp contrast to Frege's (1892) view that *sense* determines reference, since Frege identified sense with propositional meaning, and denied that a term's sense could ever be identical with its referent. (See McKinsey, 1984, 1987.)

xv. The concept of a genuine term is of course the same as Russell's (1912) concept of a 'logically proper name'. But while Russell held that (some) indexicals are genuine terms, he denied, as Frege did, that ordinary proper names are genuine terms. While Kaplan (1977) in effect showed that Russell was right about indexicals, Kripke (1972) showed that Russell was wrong about names.

xvi. I have discussed the theoretical importance of this possibility in various places. See McKinsey 1986, 1994, and 1999.

17. Points similar to these, based on the distinction between the *proposition* that $H=P$ versus the *sentence* 'H=P', occur in Tichý 1983.

18. For an application of a similar point to a defense of Jackson's (1982, 1986) argument about Mary, see Anchustegui 1997, pp. 83-84. After writing the first version of this paper, I discovered that Pavel Tichý (1983, p. 235) had raised a similar objection to Kripke's (1972) view that the sentence 'Heat is molecular motion' expresses a necessary *a posteriori* truth. Tichý claims that Kripke's view implies that the sentence 'Phosphorus is hot' says that Venus's molecules move fast. I am not myself sure that Kripke's view does imply this, but I agree with Tichý that if it does, then the view is false. I am grateful to John Turri for referring me to Tichý's paper.

xix. That this is a possibility is a consequence of the theories of names and cognitive ascriptions that I've defended in McKinsey, 1994 and 1999. See section 10 below.

20. My thanks to Mark Huston, John Turri, and Åsa Wikforss for discussions that led to the preceding two paragraphs.

21. The parenthetical qualification is necessary to rule out contradictory propositions that logically imply any proposition whatever, as well as to rule out *K*s whose existence is logically necessary and hence logically implied by any proposition whatever.

22. Thus a proposition *p* formally implies a proposition *q* just in case there is a correct deduction of *q* from *p*, a deduction whose only premises other than *p* are strict logical truths, and every step of which immediately follows from previous steps by a logically valid rule of inference.

xxiii. Though (**W**) is really the same as my proposal of 1987, p. 23, I have borrowed the useful phrase 'watery stuff in our environment' from Chalmers (1996, pp. 56-61).

24. After writing this paper, I learned that Scott Soames (2002) has recently argued at length and with great plausibility for a view of natural kind terms that is similar in several important respects to the view I'm suggesting here and that I had explained and defended in McKinsey 1987. In particular, Soames argues very persuasively (in Chapter 9) that natural kind terms are *predicates* rather than *names*. (His arguments complement the argument I had given for the same con-

clusion.) This in turn leads Soames to endorse, as I do, the view that a sentence like ‘Water is H₂O’ expresses a universal generalization, rather than a numerical identity (see Chapter 11). Finally, Soames also argues, as I do, that ‘water’ and ‘H₂O’ do not predicate the same property, but on the different grounds that if they did then (absurdly) ‘Water is H₂O’ would express the same proposition as ‘Water is water’ (pp. 276-279).

xxv. Another reason to think that Loar and Perry might endorse RQ-physicalism is the emphasis that both place on the idea that our cognitive access to sensory qualities is *demonstrative* in nature. See Loar 1997, p. 597, and Perry 2001, p. 64. I think that this emphasis on demonstrative reference to sensory qualities is misleading. According to Perry, when one says “For a feeling to be a pain is for it to be like *this*,” one would be demonstratively referring with ‘this’ to a *property*, in this case the property of being a pain (which Perry thinks is also a “physical aspect of a brain state”). But this is wrong. For a feeling to be a pain is not for that feeling to be like *the property* of being a pain. (What is it for a feeling to be like a property, anyway?) Rather, for something to be a pain is for it to be like other particular feelings, or *tokens*, of pain. So when one says “For a feeling to be a pain is for it to be like *this*,” one is using ‘*this*’ to refer demonstratively to a particular token or experience of pain and *not* to a property.

xxvi. Note that it won’t help to suppose that (P) means ‘X has Dudley and Dudley is sharp’. If (P) did mean this, then (P) would imply (D) all right. But this hypothesis about (P) makes no sense, since it makes no sense to say that the property of feeling a pain (i.e., Dudley) is sharp. At best, only *particular* pains (tokens) can be sharp.

xxvii. Even if we take seriously the proposal I’ve defended elsewhere, that the reference of a particular use of a name like ‘Descartes’ can be determined by a contingent description that is privately associated with the name by its speaker, we still don’t get the right results regarding logical possibility in Chalmers’ sense. (See McKinsey 1978a, 1978b, and 1984. Chalmers indicates (1996, p. 84), that he would take my proposal seriously.) For instance, suppose that the reference of a speaker’s use of ‘Descartes’ is fixed by a description like ‘the famous philosopher I’ve heard of named “Descartes”’. Since this description can only be satisfied by someone of whom the speaker has *heard*, it can only be satisfied in a world in which physical objects exist. So in no possible world in which the reference of ‘Descartes’ is fixed by such a description will the sentence ‘Descartes exists and there are no physical objects’ express a truth. So this sentence would not be logically possible in Chalmers’ sense, but it would actually express a proposition that is logically possible in my sense.

Chalmers has suggested to me (in correspondence) that the logical possibility of ‘Descartes is not a physical object’ can be explained on his view by the fact that the *de se* sentence ‘I am not a physical object’ can be true in a possible world W that is centered on some purely spiritual being X. (See also Chalmers 2002a, p. 612.) But this suggestion seems quite irrelevant to the issue. First, the objection I’m raising does not concern the logical possibility of the *de se* sentence ‘I am not a physical object’, but rather concerns the logical possibility of the different sentence, ‘Descartes is not a physical object’. Second, since Descartes is a physical object in every possible world in which he exists, the purely spiritual being X is of course *not* Descartes. But then I fail to

see how the possible existence of some purely spiritual being *other than* Descartes could be at all relevant to making it true that it is logically possible that *Descartes* is not a physical object.

xxviii. The two-dimensional idea in semantics ultimately derives from Kripke's (1972) concept of reference-fixing by description, as well as Kaplan's (1977) distinction between character and content. Application of the idea to modality occurs in Segerberg 1973, Stalnaker 1978 and Davies and Humberstone 1980.

29. Chalmers has pointed out to me (in correspondence) that he has in fact given an argument for the conclusion that there are possible zombie worlds. This argument assumes as premises (1) the obvious fact that zombies are conceivable and (2) the principle that conceivability entails logical possibility (in Chalmers' sense of 'logical possibility'; see Chalmers 1999 and 2002b). However, I can find no argument either in *The Conscious Mind* or in Chalmers' more recent papers to support his controversial and apparently false thesis that conceivability entails logical possibility (in his sense).

xxx. Others who express dissatisfaction with Chalmers' zombie-argument, for similar reasons, include Yablo (1999) and Perry (2001, Chapter 4).

31. For useful discussions of these matters, I am grateful to my students in Philosophy of Mind (Winter Term, 2002), especially David Stylianou, John Turri, and Matthew Zuckero, and to Erin Anchustegui, David Baggett, David Chalmers, Mark Huston, Lawrence Lombard, Lawrence Powers, Bruce Russell, Sean Stidd, and Åsa Wikforss..