Property Dualism, Phenomenal Concepts, and the Semantic Premise

Stephen L. White

Phenomenal Concepts

The property dualism argument originated in an objection to theories that entail two theses: that mental states such as pain are identical to physical states of the brain and that mental-physical identities are a posteriori. Suppose that pain is identical to C-fiber firing (CFF). Since the identity is a posteriori, a subject unaware of this fact could rationally believe what would be expressed by saying, “I am in pain, but my C-fibers are not firing.” The subject is saved from irrationality, despite believing incompatible things of the same event—the occurrence of the pain that is the firing of the C-fibers—by the fact that that event figures in the conjunctive belief under two distinct modes of presentation. In one, we take it, the pain (= the CFF) is given as a certain kind of brain state, and in the other it is given in the way pains are normally given when one has them. And suppose we think, in the first instance, of modes of presentation as aspects of the way we represent the world and not the world itself. (Call these representational modes of presentation [RMPs].) Then it seems that there must be corresponding features or properties of items in the world (in this case, features of the C-fiber firings)—by virtue of which the representational modes of presentation pick them out. (Call such features nonrepresentational modes of presentation [NMPs].) It is the burden of the property dualism argument to show that the fact or property of CFF by virtue of which it is picked out by an expression such as “my pain” (under normal circumstances) must be mental. Thus the argument purports to show that a full explanation of the a posteriori character of such mental-physical event identities presupposes a higher level mental-physical dualism (property dualism). If so, then such identity theories are incompatible with a physicalist conception of the world.

1. See Smart 1959, where he attributes the original suggestion to Max Black. See also White 1986 at 705–8 of the reprint in The Nature of Consciousness; and White 1999a, 1999c, and 2003.

Proponents of the so-called phenomenal concepts analysis of such mental-physical identities claim to answer this objection to physicalism (Loar 1990/97; Block, chap. 12, this volume). Such analyses purport to do full justice to the meaning and the a posteriori character of these identities in physicalist terms and their proponents have offered allegedly conclusive objections to the property dualism argument itself. In this section, I shall address these analyses directly and argue that there is no alternative here to property dualism. I shall then go on to present a version of the property dualism argument and to address the relevant objections.

Phenomenal concepts analyses entail that phenomenal concepts, such as the concept of being a pain, satisfy three conditions.

1. Phenomenal concepts are not equivalent to physical-functional concepts. Hence, contrary to what is believed by analytical functionalists, for example, identities such as “pains are identical with C-fiber firings” are, as alleged, a posteriori.

2. Phenomenal concepts pick out their referents directly. The relation of “pain” to pain, then, is not mediated by a mode of presentation of pain. In this it differs from the referential relation commonly thought to hold between “Hesperus” or “Phosphorus” and Venus. The ordinary assumption is that the reference, for example, of “Hesperus” to Venus is mediated by a description such as “the first heavenly body visible in the evening.” And it is assumed that “Hesperus” picks out Venus by virtue of the latter’s having the property expressed by the predicate contained in such a description—the property, in this case, of being the first heavenly body visible in the evening. To say that the referential relation of “pain” to pain is unmediated by a mode of presentation (in any ordinary sense) is not, of course, to say how phenomenal concepts do pick out their referents. On this score there are two distinct suggestions.

   a. Recognitional concept view. Phenomenal concepts are type demonstratives that pick out the properties to which they refer by virtue of the successful concept user’s capacity to recognize the relevant properties. For example, one might refer to “that taste” or to the occurrence of “that shade of red,” and one might have the capacity not only to discriminate them but to recognize them under other (e.g., counterfactual or future) circumstances. (Loar 1990/97; 600–03)

   b. Quotation concept view. Phenomenal concepts are sometimes alleged to pick out the properties they do by virtue of their “embedding” or “quoting” of instances of those properties. For example, the concept expressed by “this kind of pain” may actually embed an instance of the type of pain in question. (Block 2002: 396–98; Block, chap. 12, this volume)

3. Phenomenal concepts are not “blind.” On the recognitional concept view, we have a (second-order) phenomenal concept of what all (first-order) phenomenal properties have in common. Alternatively, we might put this by saying that we have a recognitional concept of what is common to all the things in all the extensions of the first-order phenomenal concepts (something we might call their “phenomenality”). Thus we have a concept of what it is by virtue of which phenomenal states differ from those picked out by (mere) self-directed recognitional concepts. And according to the recognitional concepts view, it is a confusion between phenomenal concepts and mere self-directed recognitional concepts that leads to the charge that
applications of the former are blind (Loar 1990/97: 603–04). Those who hold the quotation concept view, of course, already have a reply to the charge of blindness. For those who hold this version of the phenomenal concept view, in deploying such concepts we actually have (or have something similar to) an experience of the type that the concept picks out.

The question now is whether phenomenal concepts supply an adequate explanation of the a posteriori character of mental-physical identities. There is an obvious temptation to say yes, since it is a basic tenet of the theory of phenomenal concepts that they are not equivalent to any concepts in the physical-functional domain. This point could be spelled out further by saying that the two sorts of concepts have different conceptual or inferential roles. There are no relevant entailment relations between them, and, prior to the discovery of the a posteriori mental-physical identities, they are triggered by different sorts of experiences. For example, the experience that is ordinarily expressed by “I’m in pain” is triggered by pain as it is normally given to the subject of the experience. In contrast “my C-fibers are firing” is (for those who have the concept) triggered by (what is alleged to be) the same state as given through the experience of brain-scan devices and the like.

Plausible as this sounds, however, it must, it seems, be wrong. What is required to explain fully the a posteriori character of the mental-physical identities is not just that the concepts flanking the identity sign have different conceptual roles in this sense. What is required is an explanation of how the subject who claims sincerely not to believe such an identity takes the world to be. This is because the view that such identities have an a posteriori character entails that a subject could be fully rational while failing to believe or disbelieving them. Thus there must be a clear account of what the world would be like if it were the way that such an uninformed or misinformed (but still fully rational) subject took it to be. For suppose there were no such account, that every attempt to provide one led to incoherence. Then the truth of at least some of the mental-physical identities would be a priori, and the proponent of the phenomenal property approach to mental-physical identities would lack an explanation of their alleged a posteriori character.

Call the requirement that there be a coherent account of what the world would be like if it were the way such an uninformed or misinformed subject takes it to be the requirement of representational coherence. What we have just seen is that representational coherence is forced upon us by the assumption that the identities in question are a posteriori. And this requirement is stronger than Frege’s constraint as it is normally understood. Schiffer formulates Frege’s constraint as follows:

If x believes y to be F and also believes y not to be F, then there are distinct modes of presentation m and m’ such that x believes y to be F under m and disbelieves y to be F under m’. (1978: 180)

But although this is entailed by representational coherence, it does not explicitly say that the modes of presentation should be such that representational coherence is satisfied.

To see that representational coherence is what is required, notice that there is a sense in which Frege’s constraint is satisfied in the case of the person who (as we would say) believes that $27 + 17 = 44$. There is clearly a sense in which the subject has different modes of presentation of 44 associated with the two numerical expressions. The subject may recognize that it is twice 22, for example, under the mode of presentation associated with the right-hand expression, but not under the mode of presentation associated with the left-hand expression. Moreover, this subject satisfies all of the conditions that we earlier imagined the phenomenal property theorist offering as his or her account of the a posteriori character of the relevant mental-physical identities. First, the two referring expressions have different conceptual roles. That is, the subject doesn’t infer (or become disposed to assert to) sentences containing one from sentences containing the other. Second, he or she does not associate them with different evidential conditions (e.g., counting out 44 objects by counting to 44; counting out 44 objects by counting to 27, then counting to 17). But $27 + 17 = 44$ is not a posteriori. Hence what we imagined the phenomenal property theorist as offering is insufficient.

Can the phenomenal property theorist give us more by way of explanation of the a posteriori character of the mental-physical identities? In particular, can such a theorist do justice to representational coherence? It seems not. To satisfy representational coherence in the case of such alleged identities as “pain = CFF,” we need a coherent possibility to serve as the content of the belief of the misinformed subject who disbelieves all such mental-physical identities. But such a coherent possibility is just a possible world (where the notion of possibility in question is simply noncontradiction). Suppose there is such a possible world—one describable in complete detail without contradiction—at which pain is not identical to CFF or any other physical event. Alternatively, suppose there is a possible world where, though pain is a physical event, the qualitative aspect of pain is not identical with any physical property. Or suppose there is such a world where some property of the qualitative aspect of pain, such as the property of being hurtful, is not identical with any physical property. And so on. Then it is actually the case that one of these properties is not identical to any physical property, and we have the conclusion of the property dualism argument.

Suppose, then, that there is no such world at which pain and CFF come apart. Assuming that the identity is not false, it seems that we have two possibilities. It may be that there is no such coherent possibility, that the identity is, contrary to what we had assumed, a priori and that the subject who disbelieves it is irrational. But if this is not the case, then it seems that the subject hasn’t any identity or nonidentity in mind (in any ordinary sense). This amounts to a form of eliminativism. (See the discussion of local eliminativism below.) At most, it might be said that the subject who believes the identity has in mind the identity of CFF with itself. But this is something he or she could have known a priori.

The association that a subject would express by saying “CFF is this feeling,” then, would add no new fact to what the subject could have known without the experience of pain. What a subject who had lacked the experience would gain, we would have to suppose, in coming to believe that pain = CFF would be a new set of skills. This supposition would be analogous to the Lewis-Nemirov response to the Mary example (Jackson 1986, Nemirov 1990, Lewis 1988). I have no objection to such an account. Indeed, I am sympathetic to the view that what Mary gains is a set of action capacities. I believe, however, that such an analysis will be irreducibly
intentional and qualitatively and, hence, of no use to the physicalist. This is, in any case, not the line that the phenomenal concepts theorist takes (as Loar and Block make explicit), since such a theorist is committed to there being genuinely mental concepts in the identities in question. Thus I shall not attempt here to provide the argument for the irreducibly intentional and qualitative character of the skills in question.

Suppose, however, that the assumption behind this argument is denied. That is, suppose it is held that one’s belief that pain is not identical with CFF can have genuine content, even though there is no possible world that is the way one takes the actual world to be. And assume that despite the lack of such a possible world, one can be fully rational in holding it. What, though, does the content of the identity consist in? The answer for phenomenal property theorists—both those who hold that such concepts are a matter of type demonstratives backed by a recognition capacity and those who hold that they are type demonstratives backed by a more or less permanent access to the demonstratum in question—is demonstrative or direct reference. The suggestion, then, is that whatever is going on in the “CFF side” of the identity, on the “pain side” there is simply an unmediated connection between the word “pain” and its referent. Because the referent is given directly, there is no mode of presentation under which it is given and, hence, nothing on the side of the object (no property) by virtue of which the object is given under that mode of presentation. And since it seems clear that demonstrative reference gives us genuine representational content, it seems that the question of what the content of the identity is cannot pose a problem for the physicalist and phenomenal concepts theorist.

But could the need for a possible world that would rationalize the subject who believes that he or she is in pain but that his or her C-fibers are not firing really be denied in this way? It may appear so. Suppose it is said that such a subject is in a direct or demonstrative relation to the pain (i.e., the CFF) by virtue of which it is picked out by “pain” and in another relation to it (mediated by its tendency to cause certain instrument readings and the like) by virtue of which it is picked out by “CFF.” It could then be supposed that the rationality of the subject’s disbelief in the existence of the CFF is explained by the subject’s ability to imagine the pain (i.e., the CFF) without the usual (physical/theoretical) evidence of its existence. And for those who take this line it would be the existence of possible worlds at which pain is possible without the usual third-person evidence that explains the rationality of such disbelief (Boyd 1980).

This last claim, however, seems obviously false. It seems that we can imagine the pain (or the feeling of the pain, the hurtful aspect of the feeling, etc.) without the CFF itself, and not merely without its usual physical/theoretical manifestations. To see the problem for the phenomenal concepts theorist, imagine an example in which the subject has access to the pain via two demonstrative relations. The first we can take, as the phenomenal concepts theorist does, to be one in which the subject refers directly or demonstratively to “this pain.” The second, we can suppose, is one in which he or she refers directly or demonstratively (via brain scan devices or the like) to “that brain state.” Since it is the phenomenal concepts theorist’s claim that in such cases there is direct reference and no mode of presentation, he or she has no account of the rationality of belief in this case in which both references are direct. And this simply points up the fact that it is not true that demonstrative reference is accomplished without a mode of presentation of the object—a fact that Gareth Evans’s (1982) and David Austin’s (1990: chaps. 1–3) examples of the demonstrative versions of Frege’s problem make clear. Though the mode of presentation may not be (and cannot always be; see Austin 1990) linguistic-descriptive, such modes of presentation must exist and must represent the object if we are to explain the possibility of rational error in examples of this kind. In the case of both references (“this pain,” “that brain state”) our access to the CFF is via an aspect of it—just as is our access to the ship in the Evans case in which we say “that ship” while pointing out the right-hand window to the bow and “that ship” while pointing out the left-hand window to the stern. Similarly we have our representational capacities by virtue of different aspects of the objects recognized. (Many people could recognize Tony Blair by virtue of his facial features under normal circumstances, but not by virtue of any other properties—e.g., what he would look like at a costume party or made up for an amateur theatrical.)

Since both references (“this pain” and “that brain state”) presuppose modes of presentation and aspects (or properties) of the referent by virtue of which these modes of presentation pick it out, the idea (on which the phenomenal concepts view is based) of ordinary objects or events being literally their own modes of presentation cannot be made to work. The mistake seems to stem from a confusion between direct reference as it is currently understood (reference unmediated by linguistic-descriptive content) and Russellian acquaintance (Russell 1918). Were our references to our own pains direct in the Russellian sense—were they analogous to the reference, according to Russell, of the so-called logically proper names ("this,” “that”) to sense data—the referents would indeed be their own modes of presentation. But direct reference in this sense is possible precisely because sense data are thin. That is, they have no intrinsic features besides those of which the subject is immediately aware, and hence cannot figure in true, a posteriori mental-physical identities of the kind with which we are concerned. I conclude that direct reference in the current sense cannot play the role for which the phenomenal concepts theorist has it slated. Such theorist has given us no argument against representational coherence and hence no plausible alternative to the property dualist account of rational error regarding the mental-physical identities in question. I turn, then, to the positive argument that an adequate explanation of the a posteriori character of the relevant mental-physical identities presupposes the postulation of irreducibly mentalistic entities.

2. For the notion of irreducibly agential capacities see White 2004a, 2004b.

3. If it is claimed that no such instrumentally mediated state could be demonstrative and recognition, imagine that one’s brain were removed and that it controlled one’s body via radio signals. Imagine also that it was “scaled up” to the size of a neighborhood, city, or planet. We can then imagine being inside it and making demonstrative/referential references to its internal states as unproblematically as we currently do to ordinary external objects. See White 2006.
The Explanation of A Posteriori Identities

An explanation of the possibility of true a posteriori identities requires two distinct representational modes of presentation (RMPs) of the object in question. We could not, for example, represent the beliefs of a subject who (as we would say) fails to recognize that the morning star and the evening star are one and the same planet solely in terms of a singular proposition. (A singular proposition in this case would associate Venus—the planet itself—with the property of being self-identical [Schiffer 1978].) This is clear because the denial of the identity of Venus with itself would be irrational. The two RMPs in question, however, must be capable of rationalizing—that is, justifying rationally—the beliefs (as well as the intentions and actions) of a subject who believes something of the object under one mode of presentation and fails to believe it or disbelieves it under the other. And such a justification must be available at the personal level and must characterize the way the world presents itself to the subject.

We can, as we have seen, go a step further. The explanation of such an a posteriori identity requires a set of possible worlds (worlds completely describable without contradiction) that capture the content of the relevant beliefs of a subject whom we can describe as believing and as not believing or as disbelieving something of the same object. We put this by saying that what is required is a set of possible worlds that are the way the subject takes the actual world to be. And this is for three related reasons.

First, suppose that there were no such set of worlds. Suppose, on the contrary, that every attempt to describe a possible world that would rationalize and justify the subject’s beliefs brought to light a hidden contradiction. Then, far from rationalizing the subject, we would have revealed the subject’s irrationality. But, if there is no way to rationally fail to believe or to disbelieve the identity in question, then the identity is not a posteriori but a priori, and it was the possibility of true a posteriori mental-physical identities for which we were committed to providing an explanation.

Second, as I claimed above, in the absence of a characterization of the contents of the subject’s beliefs by reference to a set of possible worlds that characterize the way the subject takes the world to be, we cannot claim to have captured the contents of the beliefs at all. What we believe is something about the world and reflects a way it could be, could become, or could have been. This idea is reflected in the very notion of an intentional state’s pointing beyond itself and in the idea of an external state of affairs’ making such an intentional state true or accurate or veridical—the idea being that true or accurate intentional states are so by virtue of “truth makers” (or an appropriate counterpart in the case of accuracy).

Third, and a related point, is that the notion of a possible world (or relevant equivalent) is presupposed by the vehicle/content distinction for beliefs—that is, the distinction between what a belief represents and what does the representing. Neither the inferential role that a belief plays nor the network of connections that exist by virtue of the word-to-word connections in the subject’s language nor the functional or physical realization of such items is sufficient to give the subject’s beliefs genuine content. In the absence of connections between the subject’s words and the world, such inferential or word-to-word (or concept-to-concept) connections generate either a vicious circle or an infinite regress. In either case, they provide the subject no more than an uninterpreted formal calculus. We can put the point another way by saying that the content of belief is a matter of a condition on the world (a way the world must be if the belief is to be accurate), not merely a condition of the believer.5

The A Priori Connection between Representational and Nonrepresentational Modes of Presentation

To say that there must be a set of possible worlds that rationalize the beliefs (and the actions and intentions) of the subject who (as we would say) doesn’t believe or disbelieves the identity in question, however, raises an obvious problem. In the case of an identity such as “the morning star = the evening star,” there is no possible world with respect to which, or at which, it fails to hold for the simple reason that there is no world at which Venus is not identical with itself. What could justify what purports to be a belief to the contrary? The answer, of course, is a set of worlds at which the properties associated with two distinct RMPs are instantiated by distinct objects. The properties, for example, associated with “the last heavenly body visible in the morning” and “the first heavenly body visible in the evening,” though they are actually coinstantiated by Venus, could obviously have been instantiated by different things.

What, though, is the nature of the association between the properties and the RMPs? The set of possible worlds that justify the subject’s belief must be worlds that are the way the subject takes the actual world to be. Thus the connection between the RMP and its associated property could only be one that was given to the subject in question. Therefore it would have to be an a priori connection by virtue of the content of the RMP (for the subject)—or an a posteriori connection by virtue of the subject’s empirical beliefs. And we can eliminate the latter alternative by considering subjects who lack any relevant beliefs. In the current context this means that we can consider subjects who have no empirical beliefs that would connect pain as it is experienced directly from the first-person point of view (no mediation of brain scan devices, etc.) with any internal state characterized in neurophysiological, physical, natural, or indeed objective terms. Thus, for present purposes, we can eliminate the connection(s) via empirical beliefs. We can say that

5. It might be thought that a theory of meaning as use, such as Ryle’s, would address the need for something in addition to word-to-word or inferential connections. Ryle held that belief content was partly a matter of the actions that we are disposed to perform as a result of holding the beliefs in question. Actions, however, and intentions to act are paradigmatically intentional—states that point beyond themselves. Such a view, then, that takes intentionality to be irreducible could hardly be of use to the physicalist.
there must be an a priori connection between the subject's term (e.g., "pain"), and its associated RMP, and the property (NMP) by virtue of which that term picks out the state it does. (Call this the a priori condition.)

It follows, for example, that we could not answer the objection that a circular system or an infinite regress of word-to-word connections leaves us with only an uninterpreted infinity by appealing to bare causal relations. We could not, that is, answer the objection by postulating an inferential network realized in the functional system constituted by the brain and connected to the world via external causal relations. For these are relations that need not be accessible at the personal level, and the justificatory role that the RMPs and the NMPs are postulated to play requires that the justification be available to the subject in question. That is, the appeal to external causal connections is incompatible with the goal of rationalizing the subject with respect to the belief(s) under consideration.

The inadequacy of bare causal relations in this context can be seen clearly in light of the following examples. Imagine that on the subject of Jones's honesty, Smith is genuinely irrational. Smith believes in some contexts that Jones is fundamentally honest and trustworthy and in others that Jones is fundamentally dishonest and not to be trusted. Suppose also that Smith himself would have no difficulty recognizing and acknowledging his own inconsistency were his tendency (both to affirm and deny Jones's honesty) pointed out. We would say in this case that in the relevant sense, Smith has only one representational mode of presentation of Jones. That is, the same mode of presentation figures in both his beliefs that Jones is honest and that Jones is dishonest. (To use the file-keeping metaphor, he has, at the personal level, only one file associated with the name "Jones." In terms of my earlier formulation, then, there are not two properties associated a priori with different linguistic expressions (or different tokens of the same linguistic expression) such that those properties could have been associated with distinct objects.

This example shows, however, that the a priori condition (the requirement that the NMP be connected a priori to the corresponding RMP) is necessary for justification. For suppose that, without Smith's knowing it, he is in contact with two distinct people—Jones and his dishonest twin. And suppose that in those contexts in which Smith is disposed to believe that Jones is honest, Smith is almost always in contact with Jones, and that in those in which he is disposed to believe the contrary, he is in causal contact with Jones's dishonest double. Would this in any way undercut the assessment of Smith as irrational on the subject of Jones's dishonesty? The answer is no. As we have seen, Smith himself would describe his beliefs as irrational were the inconsistency pointed out, and (were it pointed out) he could recognize his relevant dispositions to behave as pragmatically self-defeating by his own lights. Thus it seems clear that we cannot rationalize an otherwise irrational belief set by appealing to subpersonal functional states and external causal chains to which the subject has no access.

Moreover, just as we cannot rationalize an otherwise irrational subject by appeal to such subpersonal or external considerations, so such considerations cannot turn an otherwise rational subject into an irrational one. Consider again the subject who believes something of Venus (say, that it is hot) under the RMP (associated with) "the morning star" and something incompatible (say, that it is cold) under the RMP (associated with) "the evening star." We say that such a subject is rational because although he or she believes incompatible things of the same planet, there are two routes to the referent consisting in two distinct RMPs and two appropriately related NMPs. Now imagine that we make the following scientific discovery. Astrophysicists find that the two alleged properties (being close enough to the Earth to be visible and being such as to outshine all competitors in the morning, and being so constituted in the evening) are actually both explained by a single underlying property of Venus's trajectory—say the property of being T. And imagine that being T has far greater explanatory power than any of the commonsense or theoretical properties of Venus to which we currently appeal. Suppose, finally, that on the grounds that properties must pull their weight in a causal-explanatory scheme, it is concluded that there is only one property of Venus—the property of being T—by virtue of which each of the two RMPs picks it out.

Should we conclude that in this case there is only one way in which Venus is being conceived and that the subject is therefore irrational in believing incompatible things of it? It seems clear that we should not. For it is perfectly coherent, perfectly intelligible, that the world should have been such that the last heavenly body visible in the morning was distinct from the first heavenly body visible in the evening. This is a real possibility, even if it is not a physical possibility (given what we are assuming are the physical laws and initial conditions). And we are committed to making sense of this possibility (either to explain the rationality of the subject or simply because we are committed to claims about the consequences had the laws or initial conditions been different). Thus we are committed to the existence of properties that do not pull their weight in a causal-explanatory scheme. At least this is so if a causal-explanatory theory is one that explains the physical/causal events at the actual world. Suppose, however, that this possibility that contravenes the actual laws of physics is a real one—a way the world could have been that is fully describable without contradiction. Then it is impossible to see what could be meant by saying that a subject who believes the world is this way and acts accordingly is thereby being irrational. And it seems that no empirical discovery could undermine this assessment.

Property Dualism, the Threat of Local Eliminativism

Some will likely object to the account above on grounds that rationality is misconstrued. Rationality is (they will argue) a matter of the distinctness of the subject's RMPs, not the property or properties or NMPs by virtue of which they pick out their referent. That is to say, to the extent that intentionality requires two routes to an object of which the subject believes incompatible things, the two routes are
provided by the difference in RMPs (such as the difference between “the morning star” and “the evening star”). On this view (call it the single property view), the fact that both routes go through—or pick out the object by virtue of—the same property (the property of being T) is unproblematic. Rationality, they would say, is a matter of the a priori and the a posteriori, whereas properties and possible worlds are a matter of possibility and necessity—and there is no excuse for confusing the two. (I shall discuss the modal issues surrounding the distinctions among so-called conceptual possibility, metaphysical possibility, and strong metaphysical possibility in more detail below.)

Such a stark distinction, however, with its obvious Kripkean overtones, represents, I believe, a misunderstanding of Kripke’s contribution. For consider that a priority and necessity (as they are used here) are both intimately connected with the idea of a possible world as something completely describable without contradiction. What the person believes is the case is a way that the world could (logically or conceptually) be, hence a way that the world could be that cannot be ruled out a priori. Thus it is a way that the world could be that is describable without contradiction, and therefore a way the world could be that is representable by a set of possible worlds. To suppose otherwise is, as we have seen, to confuse the vehicle of belief with the content.

Suppose we assume otherwise. Suppose it is suggested that differences in the modes of presentation required to explain the a posteriori character of the identities in question could consist in causal differences alone, unavoidable to the subject at the personal level. To accept such a suggestion (if it does not simply stem from a confusion of vehicle and content) is to adopt a position that commits one to what I shall call local eliminativism. Eliminativists regarding intentionality eschew talk of content altogether in favor of an explanation of behavior in terms drawn from the natural sciences. Similarly the local eliminativist eschews such talk in particular contexts—in this case in precisely those contexts in which an explanation of the a posteriori character of (or the possibility of rational nonbelief or disbelief in) necessary identities is required. (These contexts we might, for obvious reasons, call the Fregean contexts). Local eliminativism, however, is untenable because it is unstable. If we are committed to ascribing beliefs in such a way as to rationalize the subject (by and large), then we are committed to doing so across the board. And if we have no such commitment, then it is unclear what we could mean by the ascription of content, and we should become, regarding intentionality, eliminativists pure and simple.7 But whatever form of eliminativism is in question, it is clear that it rules out participation in the debate over the property dualism argument. For that debate turns on the requirements of our making sense of a posteriori identities. And for a proposition to be a posteriori is simply for it to be one that one could be rational in believing and rational in disbelieving or in failing to believe.

The person who claims that the a posteriori character of an identity is a matter of there being distinct representational modes of presentation of the object but not distinct properties, then, is in the grip of an overly rigid distinction between the a posteriori and the contingent. But to say this is not to address directly the question of where the objection goes wrong. The direct answer is that the proponent of the objection has no adequate account of what the distinctness of the two RMPs consists in—that is, no account that would justify the claim that the a posteriori character of the identity had been adequately explained. What, after all, could the distinctness of the RMPs consist in on such an account?

Certainly an orthographic difference—e.g., a distinction that consists in a difference in the inscription type or inscription token of the RMPs—would not be adequate. Consider the distinct inscription types “Ned,” “Block,” “Ned Block,” and “Professor Block.” Since I know one and only one person named “Ned” and one and only one person named “Block” I use these terms completely interchangeably (subtleties of style and appropriateness to the social context aside). If, then, I believe what I would express by saying something of the form “Ned is F” and what I would express by saying something of the form “Block is not F,” I could not appeal to the orthographic difference in the RMPs to rebut the charge that I am irrational. Thus an orthographic difference alone cannot provide the explanation of a posteriority that the objection in question presupposes.

The problem, in this case, is, of course, that “Ned” and “Block,” though they are orthographically different, have the same underlying causal, functional, and inferential role. Can the proponent of the single property model simply point to this difference in support of the claim that distinctness of the RMPs alone provides the explanation of the a posteriori character of the identity in question? Again there is no help for the physicalist—this time because neither causal-functional role nor inferential role (nor both taken together) are sufficient to give token utterances or inscriptions genuine representational content. As we have seen, such purely internal connections alone could never give us more than an uninterpreted calculus. And to suppose that this is all there is to our system of beliefs is to fail once again to distinguish the contents from the vehicles of those beliefs or to opt for local eliminativism.

We need, therefore, in addition to the functional or inferential role of our internal representations, some connection between those representations and the world. Could we say, then, that the combination of causal-functional role, together with an external causal connection to things in the world is sufficient? (This is perhaps the picture with which the proponent of the single property view is operating.) Again, however, the answer is no, as we saw in connection with the Smith example above. For although a system of beliefs interpreted in this way is not an uninterpreted system (in the way that a formal calculus for which we have provided no semantics is), the interpretation is not of a kind that allows us to do justice to the rationality of the subject in question. That is, we cannot, despite the difference in the functional roles of the RMPs, say how the world presents itself to the subject such that it could be rational to believe of a single object that it does and does not have some feature. We can again put the point by saying that if the proponent of the single property view is not guilty of a vehicle-content confusion, then the view collapses into local eliminativism.

7. For the constitutive role of rationality in the ascription of content, see Davidson 1973 and Lewis 1974.
The Four-Stage Argument for the Semantic Premise

What, then, is required to explain the a posteriori nature of the relevant identities? I shall provide the answer in the form of a set of requirements that make up a four-stage sequence, and I shall call the argument for the resulting final requirements the four-stage argument. The first requirement stems from the two sources we have already identified: the need the need to explain the possibility of a posteriori identities and the need to avoid the charge of local eliminativism. We have then

Requirement 1: We must say how the world presents itself to the subject who believes incompatible or contradictory things about the same object by providing a set of possible worlds that are the way that subject takes the actual world to be.

But what would such a world look like in the case of someone who believed what would be expressed by saying, "The morning star is inhabited and the evening star is not," given that there is no possible world at which the morning star and the evening star are distinct? The answer, as we have seen, is given by a world at which the object that has the individuating property associated with the first expression ("the morning star") is not identical with the object that has the individuating property associated with the second ("the evening star"). This is because, as we have also seen, we need something beyond functional states to play what we might call the rationalizing role. That is, we need something more if the RMPs are to provide a complete account of the rationality of the subject in question—the rationality of disbelief in the relevant mental-physical identities presupposed by their a posteriori status. However, two causal chains to a single object would not provide a possible world that is the way the subject takes the actual world to be. Nor could we claim, as the single property theorist does, that it is the distinctness of the expressions that explains the subject's rationality. For, as we have seen, if the distinctness is to play the justificatory role for which it is slated, it must be distinctness of content, not merely of orthography or of internal functional or inferential role. Thus we have

Requirement 2: We must satisfy the first requirement by providing two distinct properties of the object in question which correspond to the subject's RMPs and which are such that there is a possible world at which they are instantiated by different objects.

This leaves the question open, however, as to what the properties are whose instantiation by different objects at some possible worlds explains the rationality of the subject's ascription of incompatible properties to the same thing. And the answer again is relatively straightforward. As we saw in connection with the Smith example, the justification of the subject's beliefs must be available to the subject at the personal level. Thus the properties that provide the justification must be associated with the subject's RMPs in the right way. Two properties of the two distinct causal chains connecting the two RMPs to the referent, if they are properties of which the subject has no inkling, and indeed no notion, would not be associated with the RMPs in the right way. Thus there must be either an a priori association by virtue of the meanings of the RMPs or an association by virtue of the subject's other empirical beliefs. For example, if the subject believes that the last heavenly body visible in the morning is the most massive body orbiting the sun, then there would be an association of the right kind between one of the RMPs of Venus and the property of being the most massive body in a solar orbit. However, since we are free to choose a subject who has no such empirical beliefs, we have

Requirement 3: The properties (NMPs) that explain how a rational subject could fail to believe or disbelieve an a posteriori identity must be connected to the subject's RMPs a priori.

This, however, is still not sufficient. For it is plausible to suppose that there is an a priori connection between the RMP "the diameter of the Earth" and the property of being the diameter of the Earth. And it is also plausible to suppose that there is an a priori connection between an expression of the form "n meters in diameter" (where n is replaced by some particular number) and the property of being that number of meters in diameter. (These are a priori in the sense that the association of the properties in question with the RMPs depends on no empirical or a posteriori beliefs of the subject.) Suppose, however, that the diameter of the Earth = n meters at the actual world. It seems that there are two cases: (1) If the diameter of the Earth is identified with this dimension at the actual world (and n meters is understood as n times the length of the actual meter stick), then these two properties will be constantiated at every possible world. There will be no possible world that justifies the belief that the subject would express by saying, "The diameter of the Earth is n meters." Hence, as we have seen, there will be no way of capturing the content of this belief, and we will be left with local eliminativism.

Suppose, though, that we impose the requirement that properties should be thin—that is, there is nothing to such a property over and above what is understood by the subject who understands the predicate that expresses it. Then we have case (2). The thin property of being the diameter of the Earth varies from one possible world to another since one learns what the actual diameter is a posteriori and not simply by virtue of understanding the predicate "is the diameter of the Earth." Similarly one can understand "is n meters in diameter," perhaps by understanding that a meter is the length of the standard meter stick in Paris (for a nineteenth-century subject) without knowing what that length is. Hence the thin properties associated with each of the RMPs will be distinct—there will be possible worlds at which they have different extensions. And this is precisely what we need if we are to find a possible world that justifies the belief of the subject who sincerely but falsely denies that the diameter of the Earth is n meters.

Requirement 4: The properties that satisfy Requirement (3) must be thin.

If this is the case, however, we have an argument for a weakened and slightly modified version of Brian Loar's so-called Semantic Premise. According to Loar,
antiphysicalists (such as Kripke and, evidently, proponents of the property dualism argument) are committed to the following thesis.

**Semantic Premise.** A statement of property identity that links conceptually independent concepts is true only if at least one concept picks out the property it refers to by connoting a contingent property of that property. (Loar 1990/97: 600)

As I have argued elsewhere (White 2003), however, this formulation is unnecessarily strong. The (allegedly contingent) property that does the work in explaining the possibility of a posteriori identities needn’t be a first-order property of the referent in question. Such an explanation would work just as well if the property were second order or higher. We can, then, reformulate the principle as follows.

**Weakened Semantic Premise.** A statement of property identity that links conceptually independent concepts is true only if at least one concept picks out the property it refers to by connoting a contingent property of that property, a contingent property of a property of

But now notice that where Loar speaks of the contingency of the connection between at least one of the two properties connoted by the expressions flanking the identity sign and the referent of the expressions (or, in the weakened version, the referent, a property of the referent, a property of a property of the referent, or ... and so on). Requirements (1)–(4) turn on the contingency of the coconstantia of the properties (NMPs) connoted. In line with the four-stage argument, then, we have a second reformulation of the principle:

**Weakened, Modified Semantic Premise.** A statement of identity that links conceptually independent concepts is true only if the expressions flanking the identity sign pick out their referents by connoting contingently coextensive properties of that referent, or contingently coextensive properties of a property of that referent, or ... (and so on).)

What, then, are the two properties that could serve as the routes to the referent (the event which is the pain, i.e., the CFF) for the person who doubts that any "mental" event (property, etc.) is identical with anything physical and who is perfectly rational? The only property by virtue of which "pain" could pick out the CFF for such a subject would be something like the property of being painful or hurtful. Now consider any physical property $P$. Certainly, it seems, we could imagine the following: our being in a state that is hurtful without our being in a state that has $P$. But because the properties in question are thin and are connected a priori with the relevant predicates, there is no room for an illusion of contingency that is not real contingency. If so, the identity in question is false, and we have property dualism. (And if some of these conditions are not satisfied, we have no explanation of the rationality of skepticism about the relevant identities until we reach a level of higher order properties at which they are satisfied.)

**Metaphysical and Conceptual Possibilities**

A common objection to the property dualism argument is that the a posteriori character of mental-physical identities is a matter of conceptual possibilities or conceptually possible worlds, but that such worlds or possibilities need not be metaphysical possibilities. Thus, it is claimed, one cannot argue from the assumed a posteriori nature of the identity of, say, the property of being harmful and some physical property to the conclusion that there is a real possible world at which they are distinct, and thus to the conclusion that they are not identical at the actual world. Indeed, such a claim can be supported by the morning star/evening star example itself. "The morning star = the evening star" is a posteriori, and it is argued that there is, therefore, a logically or conceptually possible world at which they are distinct. It is alleged, however, that there is no metaphysically possible world at which this is the case. Thus it is suggested that the property dualism argument must ultimately rest on a non sequitur.

As applied to the property dualism argument, this is misdirected. First, the property dualism argument involves no distinction between conceptually possible and metaphysically possible worlds of the sort alleged. If “pain” and the “CFF” refer rigidly, then, as we have seen, there is no possible world describable in complete detail and without contradiction at which pain is not identical with CFF. And it is precisely this fact that motivates the property dualist’s search for the properties (or properties of properties, etc.) that are necessary for a complete and adequate explanation of the a posteriori character of what is assumed to be the identity of pain and CFF. There is, then, in the property dualism argument, nothing that would license a move from the a posteriori character of "pain = CFF" to the claim that there is a possible world at which they are distinct. Thus there is no illegitimate inference from something called conceptual possibility to some distinct notion of real possibility. There is one notion of a possible world used throughout—describability in complete detail without contradiction (keeping the meanings of our terms and the actual-world referents of our rigid terms fixed). And this notion is the one appropriate to the task of explaining the a posteriori character of the alleged mental-physical identities.

To their credit, the most prominent critics of so-called conceivability arguments (among which the property dualism argument is usually included) do not suppose that they involve such a crude mistake. Katalin Balog (1999) and Joseph Levine (2001), for example, distinguish “naive” conceivability arguments from arguments such as those of David Chalmers (1996) and Frank Jackson (1982, 1993, 1998).
(and, by extension, from the property dualism argument). And both Balog and Levine acknowledge that the proponents of the latter have no difficulty doing justice to the a posteriori necessities of the water/H$_2$O and morning star/evening star varieties. It would be a mistake, however, to assimilate the property dualism argument to conceivable arguments of even the sophisticated sort. (Balog does, as Levine does not, draw a clear distinction between them.) This is because the main arguments of Jackson and Chalmers turn on issues of modality.

In particular, they turn on a conception according to which modal truths are reducible to, or consist of, conceptual truths plus empirical facts about the referents of our referring expressions at the actual world. Such arguments can be countered by accepting brute modal truths that are not reducible in this way—for example, a brute necessary connection between a subject’s being in a certain physical brain state and being in a certain qualitative state. On such a view, then, there will be worlds that are possible in the sense of being completely describable without contradiction (even keeping our language and all the relevant actual-world references fixed) that are not possible metaphysically. That is, there will be coherently describable worlds that are not possible in the light of all the truths that are brutally necessary. Call the operative notion of necessity strong metaphysical necessity. Since there will be no inference from possibilities in my sense to strong metaphysical possibilities, does this mean that the property dualism argument begs the question against such theorists?

The answer, of course, is that the property dualism argument turns on different considerations altogether. We can see this most clearly if we simply concede (for the sake of argument) all the brute necessities that the proponent of strong metaphysical necessity desires and recognize that the property dualism argument still raises precisely the same difficulty for the physicalist. The crucial move is, as we have seen, not from the a posteriori nature of “pain = CFF” to the existence of possible worlds of any kind at which they are distinct. Rather, the move is from its a posteriori to the need for possible worlds that rationalize the subject who fails to believe or disbelieves the negation of the statement. This is just to say that if the subject is merely uninformed or misinformed and not irrational (and such a subject must be possible if the identity is a posteriori), it must be possible to give a coherent characterization of the content of the subject’s belief. But this is not to say that there is a possible world in which the belief is true. If the subject believes what would naturally be expressed in using the sentence “The morning star is not identical with the evening star” or one of its analogues, there is, as we have seen, no such world.

Where the morning star and the evening star are concerned, the answer is straightforward. Pursuing this analogy in the case of pain and CFF, however, requires that we provide a description for the CFF as it is presented to the subject who is rational but uninformed with respect to the identity of pain and CFF. But the only description could be something like “the state of mine that is hurtful.” And given that this must pick out the CFF by virtue of a thin property associated

10. The distinction between naive and sophisticated conceivable arguments is Balog’s (1999).

a priori with the expression, we have the property dualist conclusion. (The alternative, as we have seen, to supposing that the expression picks out the referent by virtue of a real property of the object—one that is thin and is connected to it a priori—is eliminationism regarding intentionality.)

Levine and Balog do, however, have a response that is relevant to the property dualism argument. Both hold a theory of direct reference according to which there need be no representational modes of presentation (RMPs) of the kind that the property dualism argument assumes must exist, and hence no corresponding nonrepresentational modes (NMPs). And certainly there is an important truth to what Levine calls nonascriptivism, which is the claim that when one uses a term, one need not have in mind, either “explicitly or implicitly, some description that would pick out its referent given a context” (Levine 2001: 53). But to think that direct reference in this sense is a response to the property dualism argument is to misconceive the issue, which is in essence Frege’s problem, and to ignore the fact that Frege’s problem arises for demonstratives as well as descriptions. As we have seen in Evans’s ship example, such purely demonstrative versions of Frege’s problem impose the same requirement as the standard examples—that we give an adequate account of the rationality, with respect to the relevant beliefs, of the subject who is rational but uninformed. And both Levine and Balog give at best short shrift to this requirement, which is the crux of the property dualism argument.

Levine, for example, writes:

When Kripke says that what’s really possible is the situation that is described had it turned out that “Water contains no hydrogen” were really true, he doesn’t mean merely that we find a possible world in which those very words express a truth no matter what they mean. It’s supposed to be that the situation thus picked out captures what we really had in mind initially by uttering the statement. So the cognitive significance of the statement must be preserved in the reinterpretation. . . . In fact, it’s precisely the use made of concepts and meaning by the advocates of the conceivable argument to which the [proponent of the theory Levine holds] objects. According to [such an] theorist, there is very little, if anything, like conceptual content, or cognitive significance, over and above the actual symbols of the relevant representations and their referents.

(2001: 53)

Doing justice to the cognitive significance of a subject’s terms, however, is not an optional feature of some solutions to the problem that motivates the property dualism argument. It is the problem which is defined by the requirement that we make sense of the rationality of the rational but uninformed subject. And what Balog and Levine offer—causal chains unavailable to the subject—are clearly inadequate. As we have seen in the Smith example, what rationalizes a subject with respect to particular beliefs is what is available to that subject, not a causal chain which is outside the subject’s ken. Levine seems to suppose that modes of presentation (RMPs) could be unavailable to the subject because he takes “meaning” to be the object of study of an empirical science—a science, presumably, of how words function in a certain community and physical environment. But this simply points up the fact that the issue is not one of meaning in this sense, but of cognitive significance—and cognitive significance for the individual whose rationality is in question. The crux of the problem is, after all, the task of explaining the rationality of the subject who is merely
uninformed or misinformed regarding pain and CFF. And this is because the task is one of explaining the a posteriori character of the identity, which is simply the possibility of doubting it while being perfectly rational.

To see that the issue is one of cognitive significance in Levine’s sense, imagine someone who is misinformed about the language of geometry and who believes that “right triangle” and “triangle” are synonyms. If such a subject believes what he or she would express by uttering the negation of the Pythagorean theorem, there is no irrationality of the kind involved in a straightforward belief in the negation of the theorem itself. What matters as regards rationality or irrationality are the modes of presentation available to the subject, not facts such as those about communal usage or causal chains to which he or she has no access.

Levine might respond that there is always a difference in the modes of presentation that are available to the subject in the relevant cases—namely the two distinct expressions (e.g., “the morning star” and “the evening star” or “pain” and “CFF”). But, as we have seen above, this proves too much. It would rationalize a completely irrational pair of beliefs—for example, beliefs that Ned Block was and was not an undergraduate physics major—simply because one was expressed using “Ned” and the other using “Block.” And this would be true even if, for the believer, the difference were merely stylistic. The objection that conceivability arguments trade on a confusion between conceptual and real possibility, then, is no threat. Either it is irrelevant to the property dualism argument, or it collapses into a direct reference claim of the kind that, in the discussion of phenomenal concepts, we have already seen reason to reject.

It is clear that in this account of the property dualism argument and of cognitive significance, the notion of thin properties does a great deal of work. Moreover, I have spoken continually of what is necessary to rationalize particular uninformed subjects in such a way that their claims to having content-laden intentional states (the uninformed beliefs) are not undermined. But I have not given a general account of what the contents of these states are. As we have seen, this is a question of what we mean when we say “what the world would be like if it were the way the subject takes it to be.” The account of the meaning of this location and of the required notion of thinness lies in the two-dimensional semantic framework that I set out originally in “Partial Character and the Language of Thought” (1982), the necessity of which is made particularly pressing by recent objections to the property dualism argument involving special modal properties. Thus it is to these objections that I now turn.

The Adam and Eve Objection

Consider the following objection to the Weakened, Modified Semantic Premise. The principle is intended to capture a necessary condition for there being a true a posteriori identity. But it could be argued that it does not. Suppose Abel is the person who originates from the union of sperm cell Adam and egg cell Eve. According to those who believe in the necessity of origins where persons are concerned, Abel could not have originated from a different sperm and egg. Thus the property of being the person who originated from sperm cell Adam is a necessary property of Abel’s—that is, one that he has at every possible world at which he exists. The same is true for the property of being the person who originated from egg cell Eve. But the identity statement

1. The person who originated from sperm cell Adam = the person who originated from egg cell Eve

is clearly a posteriori. We come to believe a statement such as this on the basis of empirical investigation, and there is no difficulty in imagining a rational subject who doubts it. We have, then, apparently, a true a posteriori identity linking concepts that connote noncontingent properties of their common referent. Hence we seem to have a straightforward counterexample to the Semantic Premise and (as I shall assume below for the sake of argument) the Weakened Semantic Premise. The example is not, however, a counterexample to the Weakened, Modified Semantic Premise because there are worlds at which the property of being the person originating from sperm cell Adam and the property of being the person originating from Eve are not constantiated—for example, worlds at which sperm cell Adam is united with an egg cell other than Eve and Eve with a sperm cell other than Adam. But now consider

2. The person who originated from sperm cell Adam at the actual world = the person who originated from egg cell Eve at the actual world.12

This is still a posteriori, and the properties connoted by the referring expressions are evidently constantiated at every possible world. Thus it seems that we have a counterexample to the Weakened, Modified Semantic Premise as well. And it would not be useful to object that the necessity of origins thesis for persons is controversial and is one that we might be tempted to deny. For whoever the truth is with regard to our own community, we can certainly imagine one in which persons are genuinely counted as the same if and only if they have the same origins in the appropriate sense.

Notice first, however, that there are two gaps in the argument against the Weakened, Modified Semantic Premise. First, there is no argument that the property of being the person who originated from sperm cell Adam is the property connoted by “the person who originated from sperm cell Adam” (and similarly for the other designating expression flanking the identity sign). Second, recall that what is being defensed is the Weakened Modified Semantic Premise. It is not required, then, that the connoted contingent properties be properties of the referent. They might, for example, be contingently constantiated second-order properties (properties of

11. I do not believe that the differences between this two-dimensional account and David Chalmers' are relevant in this context. See Chalmers 1996: 56-65.

12. This argument has been made forcefully by Ned Block (chap. 12, this volume). Block credits the second version of the example to John Hawthorne.
Phenomenal Concepts

for the Weakened, Modified Semantic Premise—which says (roughly) that if a true identity is a posteriori, there must be a possible world at which the connoted properties are not connotated—we must say that if the expressions flanking the identity sign differ in their cognitive significance, there must be a world at which the connoted properties are properties of different things. In other words, the properties must be at least thin enough to explain all the differences in the cognitive significance of the expressions that connote them. Of course, the properties connoted might be thinner than this. We might suppose, for example, that the property of having three sides and the property of having three angles are distinct, even though "has three sides" and "has three angles" do not differ in their cognitive significance (by the criterion stipulated above). However, nothing requires us to postulate properties that are thinner than is necessary to explain differences in cognitive significance, and doing so would prevent our identifying properties with functions from possible worlds to extensions at those worlds. Thus we can assume that the referring expressions differ in their cognitive significance if and only if the properties they connote are distinct.

A consequence of this is that what a definite description connotes cannot be assumed to be its meaning if meaning is assumed to be broad content. For it is arguable that in the case of the identity statement "the largest body of water on Earth = the largest body of H₂O on Earth," the definite descriptions, in addition to being coreferential, have exactly the same broad content. Indeed, from the perspective of broad content, the suggestion that the meanings of the referring expressions are constituted (in part) by the meanings of their embedded predicates, that the meanings of those predicates are a matter of the properties they express, and that they express the same properties, is difficult to avoid.

If connotation is not a matter of broad content, though, what account can we give of it? Must we give a complete account of so-called narrow content in order to provide an adequate interpretation of "connotes" in the context of the Weakened, Modified Semantic Premise?13 The answer is no. Such an account of narrow content would have to provide an account of such topics, among others, as the contents of nonreferring expressions; the contents of beliefs of hallucinating subjects; the contents of intentional states of those who, like brains in vats, are even more radically cut off from their surroundings; and so forth. For our present purposes, however, something much more modest will do. We simply need an account of the relation between representational modes of presentation, in particular (those associated with) referring expressions, and nonrepresentational modes of presentation (properties) in cases in which the referring expressions succeed in picking out a referent. And we need an account according to which the NMPs explain and constitute the relations of cognitive significance among the RMPs in accordance with the four-stage argument.14

13. For the notion of narrow content, see White 1996 and Brown 2003.
14. It is also important to repeat that cognitive significance is a matter of what is rational for individual subjects. The idea of meaning as something shared by all the language users in a community can only be an oversimplification from the perspective that takes cognitive significance as basic.
As we have seen, such an account has two components. There must be an a priori connection between the RMP and the corresponding NMP, and the NMP must be a thin property. Let us consider these requirements in turn. The first is relatively unproblematic. It is clear, as we have seen, that if the NMP is to provide the content of the corresponding RMP, the connection between the two must be a priori. But again, as we have seen, this is only half the requirement. Take the example of water. According to a well-known theory, what “water” refers to is H$_2$O. Assume that this theory is correct. That the word “water” refers to H$_2$O and that water is H$_2$O are empirical facts, known a posteriori. Hence there is no a priori connection between the predicate “is water” and the property of being H$_2$O. The property of being H$_2$O, then, is not the nonrepresentational mode of presentation that provides the route from “water” to its referent that explains the difference (for a normal subject in the cognitive significance of the representational modes of presentation associated with “water” and “H$_2$O.”

The Definition of Thinness

Consider, though, the property of being the natural kind that falls as rain, fills the lakes and oceans, and flows from faucets here (or at the actual world). It seems clear that the connection between this property and “water” is a priori for normal subjects. The connection is, after all, established on the basis of a philosophical theory, not empirical research. But it seems equally clear that the property of being the natural kind that falls as rain, fills the lakes and oceans, and flows from faucets (at the actual world) has an empirical and an a posteriori aspect.\textsuperscript{15} We can know in advance of any empirical research (i.e., know a priori) that water has this property, but we cannot know (in advance of this research) what nature water has by virtue of being the natural kind that has this property. In other words, this property confers on the quantities of matter that instantiate it an empirically discoverable essence. Thus in terms of our earlier terminology this is a thick property.

We can put this point another way and say that the property is expressed by a predicate, “is the natural kind that falls as rain, fills the lakes and oceans, and flows from faucets (here),” whose associated intention (function from possible worlds to extensions) is not invariant with respect to contexts of acquisition and/or contexts of utterance. Acquired and uttered in this world, for example, this predicate determines a function from possible worlds to extensions such that at each world a substance is part of the extension if and only if it is H$_2$O.

Consider, though, that the same predicate (in the sense of the same linguistic expression type with the same inferential role and evidential role) was acquired and used by our functional duplicates on Twin Earth. (In this paper I shall always understand Twin Earth as an alternative possible world rather than a visitable part of this world.) In such a case, the predicate would express the intension that took possible worlds into the bodies of XYZ at those worlds. Hence the predicate is noninvariant in the sense described, and the property expressed is thick.

We have, then, three distinct conceptions of thinness where properties are concerned.

1. A property is thin$_1$ if and only if it confers no empirically discoverable essence on the things in which it is instantiated.

2. A property is thin$_2$ if and only if the predicate that expresses it has an intension that is invariant with respect to contexts of acquisition and/or utterance.

3. A property is thin$_3$ if and only if the predicate that expresses it is fully intensionalized—that is, we do not have to determine the reference of the expression at the actual world in order to determine its extension at a possible world.

(Given that the equivalence of the three definitions is perhaps not completely obvious, I shall provide the relevant arguments in the appendix.)

Let me now restate the thesis I have been defending: The Weakened, Modified Semantic Premise is correct given the two conditions already stated—that the connection between the referring expressions and the properties by virtue of which they pick out their referent must be a priori and that the properties must be thin—which we are regarding as implicit in a correct understanding of the notion of “connotation.” But why should this be the case? Won’t the suspicion arise that these conditions are merely ad hoc and that even if they work in a number of cases, it is only a matter of time before we begin turning up counterexamples?

First, we should notice that the two conditions are tightly connected. Whereas the first says that the connection between the referring expressions and the properties by virtue of which they pick out their referent—or the connection between the RMPs and their corresponding NMPs—must be a priori, the second says that there must be no a posteriori component or aspect of the property. In other words, the connection must be a priori and only a priori. Second, the connection between the RMPs and the corresponding NMPs must be a priori and only a priori because together the RMPs and the NMPs have to explain how someone with no relevant empirical knowledge—someone whom we can imagine doubting any empirical proposition relevant to the referent—can succeed in picking it out. The connection between the NMPs and the corresponding RMPs must also be a priori and only a priori because the NMPs give content to the corresponding RMPs as demonstrated in the four-stage argument. And, as that analysis entails, they must give the kind of content that explains and constitutes the cognitive significance of the subject’s representational expressions and states.

Why, though, is cognitive significance the bottom line? That it is so is built into the fundamental nature of the problem that generates the property dualism.
controversy. The problem is to explain the possibility of rational error in the case of an identity that (it is assumed) is a necessary truth. Alternatively, the problem is to explain the a posteriori of the identity. But two modes of presentation differ in their cognitive significance if and only if one can be rationally justified in believing something of a referent under one and failing to believe it, or disbelieving it, of the same referent under the other. In other words, the problem that generates the controversy just is the problem of explaining a difference in the cognitive significance of two modes of presentation. 16

The proposal, then, is that the two conditions be treated as part of the meaning of “connotes” as it is used in the Weakened, Modified Semantic Premise. This is perfectly appropriate because “connotes” is a technical term related to “sense” (in Frege’s sense), and cognitive significance is (arguably) the most important aspect of that concept. And, as the foregoing suggests, this is the meaning of “connotation” that Loar needs if the Weakened, Modified Semantic Premise is to serve its purpose. Moreover, the thick/thin distinction is a stand-in for a full-blown theory of narrow content, where narrow content is the content that constitutes and explains cognitive significance and resolves the Frege puzzles. Hence it is not unmotivated to understand connotation as narrow content and to say that the Weakened, Modified Semantic Premise is correct where connotation is understood in terms of narrow rather than broad content. (This is, of course, subject to the qualification presented above that no full-blown theory of narrow content will be presented and that none is necessary in this context.) In any case, nothing turns on the identification of connotation and narrow content. We could easily restate the Weakened, Modified Semantic Premise explicitly in terms of the thick/thin distinction and avoid the use of “connotation” altogether.

Explaining the Illusion of Contingency

Even if it is clear that the “two conditions” are well motivated, we can still ask how the thick/thin distinction resolves the problem that (1) genuine identities are necessary, (2) the identities in question are a posteriori, and (3) there must be possible worlds that rationalize the belief of the subject who doubts the identity, and (4) there must be NMPs that give the RMPs their contents. We can sharpen this question if we imagine a skeptical and unsympathetic objection. “You are committed,” the objector might say, “on the one hand to the idea that genuine identities are necessary—hence to the idea that there is no possible world, for example, at which pain is not the same thing as C-fiber firing (CF). On the other hand,” the objector might continue, “You are committed to there being some connection between things on the side of rationally justified belief and RMPs and things on the side of properties, possible worlds, and NMPs, since you believe that the latter are necessary to give the former their content. In particular, you are committed to there being a set of possible worlds that rationalize the belief of the person who doubts or disbelieves the identity of pain and CF. Of course, the broad/narrow or thick/thin distinctions show that these commitments are not immediately inconsistent. But you have not yet answered the question of what the possible worlds are that rationalize the subject’s beliefs in the case of these sorts of identities." And we can make this objection clearer if we consider a possible reply that we might make following Kripke and the objections to such a reply.

The objector challenges us to satisfy the following desiderata. The first is to say which propositions they are whose real contingency explains the apparent contingency of the necessary a posteriori identities in question. The second is to say how they are related to the necessary a posteriori identities such that a fully rational subject could be justified in doubting the latter. The first desideratum amounts to the one spelled out in the four-stage argument that there be possible worlds that capture the way the subject takes the actual world to be. And the second adds to that the requirement that there be a general characterization of the way that such worlds are related to those that constitute the truth conditions of the identities in question. Moreover, in line with the four-stage argument, these worlds that establish the genuine contingency of the proposition that justifies the subject’s doubting the a posteriori identity must, in some appropriate sense, provide the content of that identity.

Consider the two replies that, on the basis of Kripke’s discussion, we could make to such a challenge. First, there is the suggestion that what justifies the subject’s disbelief in necessary identities of this kind is the existence of a world that is, in a qualitative sense, epistemically the same as the actual world and with respect to which a “qualitatively analogous statement” is false. For example, according to Kripke, “Heat = molecular motion” is (if, as we may assume, true) necessarily true. But its a posteriori character, apparent contingency, or the illusion of its contingency is explained by the existence of a possible world at which the sensation of heat is produced by something other than molecular motion. And Kripke evidently means that with regard to such a situation, the statement “The phenomenon that produces the sensation of heat is not molecular motion” is true (where by the “sensation of heat” means the sensation produced by heat at the actual world), even though “Heat is molecular motion” also is true at this possible world, as at every other (1972: 140-54).

Viewed as a necessary and sufficient condition (or even a necessary or sufficient condition) of our having an explanation of the illusion of contingency of the identity in question (or of our having a rational justification of the subject who fails to believe the identity), however—and it is unclear how Kripke intended it—this suggestion is open to a number of objections. First, there seems to be no account of qualitative similarity that doesn’t raise serious difficulties. If, for example, it means equivalence as regards sense data, then it relies on an analysis of perception and experience that is open to apparently conclusive objections. (I shall consider this interpretation shortly.) This is also a problem with taking the account as providing a necessary condition (or having an explanation of the apparent contingency of the identity in question or a justification of the subject who disbelieves it.)

16. To say that there are two modes of presentation is not to beg the question against the view that the referent might be its own mode of presentation, as long as it has another mode of presentation that is distinct. A view that denies this is a nonstarter in this context.

17. Or seems to doubt. I shall ignore this qualification in what follows.
Second, there is the Boyd objection considered above (Boyd 1980: 83–85). If the account is taken as a sufficient condition of our having the explanation and justification in question, then, as Boyd points out, it won’t serve Kripke’s purpose where pain and C-fiber firing are concerned. Kripke’s argument is that in the case of the alleged identity of pain and C-fiber firing, the apparent possibility of our having pain without CFF and vice versa cannot be explained by citing the real possibility of our having the sensation of pain without CFF (by analogy with heat). This is because, as Kripke points out, the sensation of pain is pain. In other words, according to Kripke, a possible world at which we have the sensation of pain without CFF is one at which we have pain without CFF. Hence, according to Kripke, the attempt to explain away the apparent contingency of “pain = CFF” entails its real contingency, hence its falsity. Thus the explanation fails. According to Boyd, however, even if we assume that “pain = CFF” is true, there are (pace Kripke) possible worlds at which qualitative analogues of the identity are false. For there are worlds at which we have pain (and hence, by hypothesis, CFF) without any of the sensations normally associated with CFF given as such—that is, as a brain state. There are, for example, obviously possible worlds at which all of our instruments give us the sensory experience associated with the absence of any CFF—they may simply not be working. Conversely, we could have the qualitative experiences associated with CFF (given as such) without the sensation of pain, because we could have the former without CFF (i.e., pain). Thus, according to Boyd, Kripke’s argument for dualism can be blocked.

There is a third objection to this explanation of apparent contingency. As we saw above, the set of possible worlds that rationally justify the subject who disbelieves a necessary (but a posteriori) identity do so because they are the way the subject takes the actual world to be. They capture the content of the subject’s belief (in the sense of the way the world presents itself to the subject), though not the truth conditions of the belief (since if pain is CFF, there is no world at which they are distinct). If this is the case, however, then pointing to a world that is qualitatively like the actual world (but at which a qualitative analogue of the identity in question is false) is not an appropriate way of rationalizing the subject and explaining the illusion of contingency. For it implies that the content of the subject’s belief can be captured in sensational or sense-datum terms, and this seems clearly false. If one believes that water is not H₂O and that there can be one without the other, this seems radically different from the belief that one could have the sensations commonly associated with one without those generally associated with the other.

Kripke never commits himself to an interpretation of “epistemic identity in the qualitative sense” in terms of sensations or sense data. What if we consider more liberal interpretations? Suppose, for example, that we construe epistemic equivalence in the qualitative sense as either observational equivalence in one of the many senses of “observational” or, even more vaguely, “evidential” equivalence. Would this help to address the three objections that we have just seen to what I shall call the qualitative equivalence criterion? Certainly it would address the first, that talk of sense data is objectionable per se. It seems less useful, however, in addressing the Boyd objection and the objection that a qualitatively specified condition cannot provide the content of the belief of the subject who doubts the pain-CFF identity. To the extent that it does address these objections, it seems to presuppose an instrumentalistic and phenomenalistic view of science and of content that few are likely to find attractive. These objections cannot be made conclusively against the more liberal interpretation of qualitative equivalence, however, because of the vagueness of qualitative equivalence and because we cannot pursue here such topics as phenomenalism, instrumentalism, and realism in the philosophy of science. Thus I shall sketch an independent problem.

Let us recall that the point of the reference to worlds qualitatively equivalent to the actual one is to explain the apparent contingency of necessary identities, and, more generally, to rationalize (i.e., rationally justify) the subject who doubts such an identity. Even more generally, it is part of the project of providing a rational justification of (intuitively) rational error wherever it occurs. Now consider a (badly misinformed) subject who believes that water is not H₂O. Following Kripke we say that his or her belief is rationalized by a possible world at which, say, the most plentiful colorless, odorless, life-sustaining liquid is not H₂O. But now consider “colorless,” “odorless,” and “life-sustaining.” Suppose that these, like “water,” are themselves natural kind terms. Then they themselves are capable of generating exactly the same kind of apparent irrationality. That is, they are capable of generating the ostensible irrationality associated with Fregian problems in which a subject doubts what he or she would express by uttering a statement that is true with regard to every possible world. If, for example, there is a discoverable empirical essence to being colorless, or odorless, or life-sustaining, then a rational subject could believe that a liquid was colorless but that it lacked the physical property with which the property of colorlessness was identical. In this case the subject would doubt a necessary truth, and we would, as yet, have no rational justification of his or her beliefs. And if we try to rule out this possibility by appealing to properties like looking colorless, we are back to the original phenomenalistic or sense-datum interpretation of qualitative equivalence. (And again recall that even if we don’t treat so-called qualitative terms as natural kind terms, we can easily imagine rational subjects who do.)

I now turn, then, to the second way in which we might, following Kripke, say which propositions are whose real contingency explains the apparent contingency of the necessary a posteriori identities in question. And this, of course, is to say how they are related to the latter such that a fully rational subject could be justified in doubting them. Kripke says:

In the case of identities between two rigid designators, the strategy [involving the appeal to the sameness of the subject’s epistemic position, qualitatively speaking] can be approximated by a simpler one: Consider how the references of the designators are determined; if these coincide only contingently, it is this fact which gives the original statement its illusion of contingency. (1972: 150)

Unfortunately, this does not necessarily address the problems with the qualitative equivalence criterion (for example, that the representational modes of presentation that determine the references of the designators may themselves raise Fregian
The Dilemma for Kripke

We can, then, think of Kripke as facing a dilemma: On the one hand, Kripke can embrace a highly reductive account of the source of the illusion of contingency. That is, he can explain the illusion that water could fail to be H₂O, say, by pointing to possible worlds at which something that produces the same sensory experiences or sense data that water does (at the actual world) is not H₂O—that is, worlds at which the substance that looks colorless, has no odor, and so on is not water. This, as we have seen, is open to two objections:

1. This doesn't seem to capture what a subject means (or has to mean) in claiming that water is not H₂O. Such a subject might explicitly claim that he or she does not merely mean that something that produces the same sense data as water is not H₂O, but that water is not H₂O.

2. Moreover, this is open to the Boyd objection that if this is an adequate account in the case of water and H₂O, it can be applied to the case of pain and C-fiber firing to show how the illusion of contingency may be explained without denying the identity of pain and CFF. We can explain the illusion in a way exactly analogous to the heat case by pointing to the existence of worlds at which the sense data associated with CFF given as such (i.e., as a brain state) are produced by something other than pain (and hence other than CFF) and worlds at which CFF does not produce the sense data associated with CFF (given as such) at the actual world.

If, on the other hand, Kripke refuses to adopt a radically reductionistic account of the illusion of contingency, then it seems he will face an infinite regress of descriptions. If he refuses to embrace the phenomenalist strategy, then it seems he will have to explain the illusion of contingency by appeal to descriptions by virtue of which the designators in question pick out their objects. Suppose, however, that these descriptions themselves involve natural kind terms that generate further illusions of contingency—e.g., terms like "is a liquid," understood as a natural kind term, as explained above. (And recall that even if in fact these terms are not used as natural kind terms, we can easily imagine communities in which they are.) Suppose, then, that being a liquid is itself identical to a physical property and that this identity is a posteriori. Then explaining the difference in the cognitive significance of "water" and "H₂O" for a normal subject in terms of the difference in the associated descriptions for that subject will lead to another set of beliefs whose obvious rationality has no explanation on Kripke's account. And it seems clear that we cannot have an infinite backward regress of such descriptive accounts. If so, then Kripke's theory will provide no general account of the rationality of error in the case of a posteriori identities.

Gap-Inducing Linguistic Devices

What is required to resolve this dilemma for Kripke is a more explicit characterization of the general account of rational error that we have already seen in outline—that is, the Weakened, Modified Semantic Premise, together with the two implicit conditions: that the relation between the RMPs and their corresponding
NMPs must be a priori and that the properties that play the role of NMPs must be thin.

The problem with what we have seen so far concerns the notion of thinness. Thin properties were originally characterized as those that conferred no hidden (or empirically discoverable) essence on the things that instantiate them. And although the application of this notion in the context of natural kind terms seems straightforward, its application in the context of the Adam and Eve example does not. To be sure, Kripke’s account of the necessity of origin where persons are concerned has certain obvious analogies with the widely accepted understanding of natural kinds. But rather than explore these analogies to expand the application of the concept of the lack of a hidden (or empirically discoverable) essence, it seems preferable that we should try to deploy a broader and more abstract conception of thinness.

It is the need for a more abstract notion of thinness that motivates the two alternatives to the characterization in terms of the absence of a discoverable essence. According to the first alternative, thin properties are those expressed by predicates whose intensions are invariant across contexts of acquisition and utterance. According to the second, thin properties are those expressed by predicates that are fully intensionalized in the following sense: given a complete description of a possible world, we can tell what the extension of the predicate is at that world without first having to answer empirical questions about the actual world. And as we have seen, properly understood, these three characterizations of thinness are equivalent.

We can, in fact, take this line of thought further. We can appeal to the notion of a predicate that is not fully intensionalized in order to define what we might call gap-inducing linguistic devices. Such devices take fully intensionalized predicates into predicates that are not fully intensionalized. Consider the phrase “at the actual world.” Assume (as is the case by almost anyone’s lights) that “is square” is fully intensionalized. (Given a possible world PW, we can settle the question of which things are square at that world by reference to the facts at that world alone and without knowing anything about the actual world.) Contrast this with the predicate “is square at the actual world.” To know which things instantiate the property that this predicate expresses with regard to PW, we have to know what shape they have at the actual world.19 “At the actual world,” then, is a gap-inducing linguistic device.

Similarly, it would be widely accepted that proper names are such devices. Suppose we want to know with respect to a certain possible world PW whether the following scenario obtains: that, in addition to being called “Mark Antony,” Julius Caesar did all the things associated with Mark Antony at the actual world and none of the things associated at the actual world with Caesar (i.e., Caesar played the Mark Antony role at PW), and that Mark Antony played the Caesar role. To answer this question, we need to know who Caesar and Mark Antony are at PW. And to do this,

---

19. We have to stipulate, say, that if they don’t exist at the actual world, they aren’t square at the actual world.

we have to determine whom “Caesar” refers to, and whom “Mark Antony” refers to, at the actual world (arguably by tracing the relevant causal chains at the actual world to their sources) and then determine what those persons did at PW. In other words, we cannot settle the issue of who Caesar and Mark Antony are at PW on the basis of what we have access to as users of the terms “Caesar” and “Mark Antony” and a qualitative description of PW. Other plausible examples of gap-inducing devices include demonstratives and indexicals.

Something analogous might be said about natural kind terms. A plausible characterization of the content of a particular natural kind term would be something of the form

The natural kind that realizes D at the actual world

where D stands in for a description couched in terms of predicates that express the macro-level properties to which we as a community or we as individuals have access. (Thus we might think of natural kind terms as containing implicitly the gap-inducing device “at the actual world.”)

If this account is plausible, it suggests a general strategy for explaining rational error, a strategy that has a clear application in the Adam and Eve case. We can, as we have seen, “strip off” such apparently pure rigidification devices as “at the actual world” (devices that do not themselves contribute descriptive content to the predicate in question) while “filling in” the gaps induced by such devices as proper names and natural kind terms with descriptive content available to the subject. But, as has been implicit in what has already been said, these two strategies, even in combination, are not sufficient. Neither alone nor in combination do they deal with the problem we saw in connection with Kripke—that we cannot rule out an infinite regress of natural kind terms. We cannot rule out, for example, the possibility that every descriptive content to which we appeal to fill in a gap induced by a term like “water” (terms such as “colorless” and “wet”) will themselves be natural kind terms. And if so, there will be no descriptive, fully intensionalized vocabulary available to the subject by appeal to which all such terms could be eliminated.

There is another strategy, however, by which we can abstract from any empirically discoverable essence implicit in the predicates to which our empirical situation limits us. Consider the common characterization (to which I have already alluded) of the content of a predicate (and hence of the property expressed) in terms of an intension—that is, a function from possible worlds to the extensions of that predicate at those worlds. The intuition behind the equation of such a function, the content of a predicate, and the property expressed by that predicate21 is the same as the intuition that equates a proposition with a set of possible worlds. In the propositional case, the content of the proposition is whatever condition holds for all the worlds in the set. By analogy, the property expressed by the predicate is what

20. But isn’t there a hidden essence involved even if we drop this phrase? For each possible world, it is an empirical question which natural kind it is that satisfies the description. This, however, is as it should be. The case is analogous to “the murderer of Smith” understood nonrigidly.

21. Or alternatively, the reduction of the property and the content to the function or the representation of the content and the property by the function.
the elements of all the extensions at all the possible worlds have in common. The fact that we look at the extensions of the predicates at all logically possible worlds eliminates the sort of problem raised by such predicates as “creature with a heart” and “creature with a liver.” Because these allegedly have the same extension at the actual world and yet (intuitively) differ in their contents and in the property expressed, we cannot identify their contents with their actual extensions. This problem is solved, however, by abstracting from the contingencies at the actual world. It is precisely because we can imagine a world at which creatures who in their natural, normal, and healthy state have a heart and no liver (and vice versa) that the two predicates differ in their cognitive significance and thus in their content. Thus the criterion of the identity of a property that makes it what is common to all the members of the extensions across all possible worlds is an attractive one.

The General Account of Apparent Contingency

The identification of the property connoted by a subject’s linguistic expression with a set of extensions across all possible worlds does not itself provide the needed account of apparent contingency, of course. Rather, it provides the model for that account insofar as it involves the use of possible worlds to abstract from actual world contingencies. The difficulty, as we have seen, is that such expressions as “at the actual world,” in contexts such as “is the natural kind that realizes the water role at the actual world,” inject actual-world contingencies not only into the actual extensions of predicates but into their extensions across possible worlds. Whereas “the natural kind that realizes the water role,” understood nonrigidly, or in a fully conceptualized way, has an extension that varies across possible worlds (it is $H_2O$ at the actual world, XYZ at Twin Earth, etc.), “the natural kind that realizes the water role at the actual world” picks out the same natural kind at every world. Thus a contingent and empirical fact about the actual world again prevents the extension of a predicate—this time across possible worlds—from corresponding to what is, by the criterion of cognitive significance, the content of the expression in question.

There is a solution in this case, however, and it is analogous to the solution in the earlier case. We can arrive at an extension of a predicate that reflects its content if we can abstract from the actual world contingencies that break the connection between the content of a predicate and its extension across possible worlds. Furthermore, we have the tools at hand to do so. What I have called the partial character representation of content calls for a two-dimensional matrix: rows correspond to possible worlds construed as contexts of acquisition, and columns to worlds construed as contexts of evaluation. Now consider a fully intensionalized predicate (assume, for the sake of argument, that “is $H_2O$” is an example). Because the predicate is fully intensionalized, its possible-world extensions do not vary as a function of its context of acquisition (see table 11.1). That is to say, reading down each column we have the same sequence of entries as in every other column. (Equivalently, reading across each row, the entry is always the same.)

Thus for fully intensionalized predicates, the matrix representation for their content reduces simply to the function from possible worlds (construed as contexts of evaluation) to extensions—that is, just the kind of function from possible worlds to extensions that has traditionally been thought to provide the contents of predicates. And for fully intensionalized predicates, this is a completely adequate characterization of their contents.

Contrast this with the case of a predicate such as “the natural kind that plays the water role at the actual world.” Acquired or uttered at the actual world, this has $H_2O$ as its extension at every possible world (construed as a context of evaluation). Acquired and/or uttered on Twin Earth, however (construed as a possible world and a context of acquisition and utterance), it has as its extension XYZ at every possible world (construed as a context of evaluation) (see table 11.2).

We might put this by saying that contingencies at the context of acquisition or utterance are leveraged into necessities by our method of evaluating our subject’s beliefs and utterances at possible worlds. Instead of carrying a fully intensionalized content to the various contexts of evaluation, we carry the thing that in fact satisfies the content (in this case a description) at the context of acquisition or utterance. We can say, then, that the function of the possible-world apparatus in bringing extensions in line with content (understood in terms of cognitive significance) that we saw earlier in connection with “is a creature with a heart” is undermined by our method of evaluating our expressions at the possible contexts of evaluation. And, of course, everything that has been said of the predicate “is the natural kind that plays the water role at the actual world” applies to the predicate “is a creature with a heart.”

---

22. See White 1982 for the details regarding what is kept fixed in the consideration of worlds as alternative possible contexts of acquisition of a term. (The relevant internal, functional psychologies of subjects must be roughly the same or close enough.)
the water role at the actual world" has an analogue with respect to "is the person originating from sperm cell Adam at the actual world."

But how does this treatment of possible worlds as possible contexts of acquisition help with our problem—indeed, Kripke’s problem—that there may be no vocabulary available to the subject in which to express a fully intensionalized content available to the subject that has the same cognitive significance for the subject as the relevant predicate?

The answer is that there is a move open to us that is analogous to the move from taking extensions of predicates at the actual world to taking their extensions across all possible worlds in order to fix their contents. As we have seen, we can abstract from the contingency that at the actual world "is a creature with a heart" and "is a creature with a liver" have the same extensions by looking at their extensions at all possible worlds. Similarly, we can abstract from the contingency that it is H₂O that realizes the water role at the actual world by looking at its extensions across all possible contexts of acquisition and utterance. In other words, we explain the property that explains the possibility of rational error in a case of this kind not as the property expressed by some (set of) descriptive predicates available to the subject, but as one characterized by, reducible to, or identical with, the two-dimensional matrix that abstracts from the contingencies of the actual world not only as it is construed as a context of evaluation, but as it is construed as a context of acquisition and/or utterance.

By identifying the thin properties needed to make the Weakened, Modified Semantic Premise immune to counterexamples with the two-dimensional matrices associated with the predicates in question, we can overcome the dilemma confronting Kripke’s treatment of this issue. And although Boyd’s strategy for dealing with the so-called illusion of contingency in the case of pain and C-fiber firing raises a problem for Kripke, it presents no problem for the present account. Nor is there any question about whether the properties this account yields are thin enough. Keeping fixed the content (in the sense of cognitive significance) that individuals associate with the referring expressions that occur in the kinds of identities in question, we let everything else vary, compatible with consistency. If, in so doing, we produce a description of a world at which the identity, so understood, fails, then we have a rationalizing explanation and justification of the subject who calls that identity into doubt.

Does this reference to cognitive significance render the account circular? After all, one of our stated goals was to explain cognitive significance. I think the answer is no. In "Partial Character and the Language of Thought" (1982), I argued that what was held constant across possible contexts of acquisition was the functional makeup underlying the subject’s use of the term in question, and this principle seems sufficient for our purposes here. For the physicalist is committed to a sense of meaning that supervenes on what is in the head—namely, the sense in which meaning just is cognitive significance. This is simply the upshot of examples we have already seen. (Recall the example of Smith, whose irrationality regarding Jones’s honesty was independent of the facts regarding external causal chains precisely because the equivalence in the cognitive significance of his referring expressions was a matter of what was available to him at the personal level.) If this is the case, however, then the physicalist cannot object to a reductive, internal characterization of cognitive significance. (If identity or sufficient similarity of the relevant functional states is not enough, we can always add the requirement that the realizations of the functional states be “of the right physical kind.”)

We can, then, appeal to the two-dimensional framework to answer the objection to Kripke regarding natural kind terms—and to provide enough of what would be provided by a positive theory of narrow content to deal with the objections to the property dualism argument and the Weakened, Modified Semantic Premise. And if we can dismiss the objections on grounds that the physicalist cannot dismiss, then the dilemma for physicalism stands: either dualism regarding sensations such as pain or properties of such sensations on the one hand, or eliminativism regarding the intentional on the other.

But note: Not only is the property dualism argument not intended to provide an account of narrow content—it is not intended to provide a positive account of any sort. It is an argument that the physicalist account in question faces an insurmountable obstacle. Thus it has been a matter of working within the physicalist framework to show that it cannot do justice to the concepts that it purports to explicate. But working within such a framework and making the minimum number of changes is not likely to be the best way of constructing a positive alternative. My own current preferred account of reference, for example, would give pride of place not to bare causal chains, but to irreducibly agential skills and capacities. (See the opening section of this chapter.) Such an account would require that what we keep fixed as we move from one possible context of acquisition to another are just those skills and capacities that underlie the agent’s use of the relevant term. Having noted that what I have here is a negative argument and not a positive account, I can legitimately defer this discussion.

The Weakened, Modified Semantic Premise stands, then, and this account makes it clear why. The logical possibility that defines the possible worlds to which we appeal and that provides the metaphysical conclusion to the property dualism argument and the real conceivability that provides the premise are at bottom the same: describability in complete detail without contradiction. The simplicity of this connection is of course complicated significantly by the broad content that—in the form of gap-inducing linguistic devices—governs the evaluation of our utterances and beliefs at other possible worlds. It is the fundamental need to explain the possibility of rational error, however, that provides both the need for a notion of content that tracks cognitive significance and the notion that reestablishes the fundamental connection between what we can rationally conceive and what we are committed to regarding as possible in the most fundamental sense of the term.

Appendix

**Thick₁ → Thick₂**

Suppose a property \( P \) is thick, then there is some feature \( F \), discoverable at the actual world, such that for any possible world, something is \( P \) if and only if it is \( F \). But if this is how the predicate that expresses \( P \) works, then had the actual world been different, \( P \) would have been different. And this entails that the predicate does not have an intension that is invariant with respect to contexts of acquisition and/or utterance.

**Thin₁ → Thin₂**

Suppose there is no empirically discoverable essence. Then either (a) there is no empirical content as in, say, mathematical predicates, logical predicates, automata theoretic predicates, and so forth. Alternatively, (b) it varies from world to world. If (a), then it doesn't matter where we acquire the term; we can simply look at a possible world and tell whether something satisfies the predicate at that world.

But the same is true if (b). Consider again the description "the natural kind that falls as rain, fills the lakes and oceans, and flows from the faucets." And suppose that the description gives the content of "water" and is understood as nonrigid. Then although at each possible world there is much to be discovered empirically about the referent of the description, there is nothing empirically discoverable that is true across all possible worlds. (All that is true across all possible worlds is that at each world where water exists it falls as rain, etc., and we know this a priori.) Hence there is no empirically discoverable essence. But in this case it doesn't matter where we acquire the term or where we utter it, it has the same intension—that is the same function from possible worlds to extensions.

**Thin₃ → Thin₁**

If a predicate has no devices of direct reference and is fully intensionalized, then there is nothing which is context sensitive. We can apply it to a possible world without having to know anything about the references of any referential devices at the actual world (besides what we know on the basis of an understanding of the language alone). Thus there is no room for a hidden or empirically discoverable essence. We could put this metaphorically by saying that there is nothing we have to "carry to another possible world," along with the predicate, to determine its extension at that world.

**Thick₃ → Thick₁**

Suppose the predicate is not fully intensionalized. How, then, could we apply it to a possible world? We don't have a fully determinate content to "carry" to that world. What is the alternative? We get the necessary and sufficient conditions that we need by determining the referent at the actual world and taking the necessary and sufficient conditions for being that thing across possible worlds. But this gives us an empirically discoverable essence.

ACKNOWLEDGMENTS

I am grateful to Torin Alter, Ned Block, and David Chalmers for comments and suggestions regarding earlier drafts.

REFERENCES


T W E L V E

Max Black’s Object to Mind-Body Identity

Ned Block

In his famous article advocating mind-body identity, Max Black (1966) put forth an objection that he thought was first put to him by J. D. Trout, a prominent figure in philosophy of mind. Beck's (1986) "Knowledge Argument," as it is often called, turns on the following assumption: if a being has a property instantiation of a property that is not possessed by any other being, then that being is identical with the other being. This assumption is itself an objection, because it assumes that two beings can be identical if and only if they have the same properties.

I will say a bit about what the basic idea is and compare it with the Knowledge Argument. I will also examine John Perry's (2001) book, which rigorously develops the knowledge argument as well as some variations. I will then return to Smart's (1986) essay on the topic and arguments inspired by it.

1. Stephen White (1986; 2006, chap. 11, this volume) defends the property dualism argument. John Perry's (2001) book on dualism is a response to Jackson's Knowledge Argument. In his 1993 paper, "The Self," White argues that Jackson's argument is flawed because he assumes that two beings can be identical if and only if they have the same properties.

2. I discovered rather late in writing this chapter (or even earlier) those in the chapter's last third, are somewhat related to the ideas of Descartes. As I understand it, Arnaudic's (1986) critique of the mental substance by arguing that nothing objective "backsides" to the mental whose objective dualism is indebted to Tyler Burge for drawing the Rezaom