

Materialism without Reductionism: What Physicalism Does Not Entail

Richard Boyd

IN A RECENT WORK of substantial importance, Saul A. Kripke (1972) offers an alternative to the received accounts of reference, necessity, and essential properties. In a small section of the paper Kripke applies his account of necessity to certain traditional "essentialist" objections to mind-body identity—objections according to which mind and body cannot be identical because they have different essential properties. According to Kripke, standard materialist rebuttals to these

objections rest on a mistaken account of essential properties. He suggests that, in fact, no rebuttal is possible.

In this paper I hope to accomplish two things. First, I intend to show that Kripke's discussion of reference and necessity constitutes a significant contribution to our understanding of the mind-body problem, not only because his account explicates better certain objections to materialism but also because something like Kripke's account of reference is required for a satisfactory defense of materialism. Second, I intend to show that the particular essentialist arguments Kripke directs against materialist theories of mind are, though ingenious, entirely unsuccessful.

1. "Essentialist" Objections to Materialism

A striking thing about materialist solutions to the mind-body problem is the strong and conflicting philosophical intuitions they seem to elicit. On the one hand, it has seemed to a great many philosophers and scientists that the doctrine that mental phenomena are really a species of the physical is an almost unavoidable conclusion in the light of the increasing success with which physical scientists

Copyright © 1980 by Richard Boyd. An early version of this paper was presented to the Philosophy Discussion Club of the Sage School of Philosophy at Cornell University. I am grateful for many helpful comments on that version, especially for the detailed comments of Carl Ginet and Norman Malcolm. I am indebted to a large number of other colleagues for helpful criticism and discussion of the positions defended here. I especially want to thank Rogers Albritton, Ned Block, Oswaldo Chateaubriand, Alex Goldstein, Barbara Koslowski, Norman Kretzmann, Richard Miller, Hilary Putnam, Sydney Shoemaker, Robert Stalnaker, Nicholas Sturgeon, and William Wimsatt. [*Editor's note:* This paper was solicited as a discussion of nonreductionist materialism and its relation to "essentialist" critiques of mind-body identity theses. The author kindly agreed to expand the paper to include an introduction to the central linguistic issues raised by materialism.]

have explained complex biological phenomena. It seems to them overwhelmingly unlikely that physical explanations for features of our mental life will not also be forthcoming.¹ On the other hand, many of the same thinkers, some of the time, and a great many others, all of the time, share the contrary intuition that it is absurd and incoherent (or, perhaps, even mad) to assert that mental phenomena are physical. Mental phenomena are simply the wrong *kind* of thing to be physical; they are *essentially* nonphysical.

At least since Descartes, such intuitions have been understood in terms of a putative contrast between the essential properties of mental and physical phenomena. Mental phenomena are said to have as *essential* properties certain properties (like privacy or introspectability) that are not possessed or, at any rate, are not possessed *essentially* by physical phenomena; alternatively, physical phenomena are said to have certain *essential* properties (like spatial location or publicity) that are not *essential* properties of mental phenomena. It is sometimes maintained that mental and physical phenomena have contradictory essential properties (that physical events are essentially spatial, for instance, while mental events are essentially nonspatial). In all these cases, anti-materialist intuitions are understood as intuitions that the properties that are logically possible (logically necessary, logically impossible) for mental phenomena are different from those properties that are logically possible (logically necessary, logically impossible) for physical phenomena, and therefore that mental phenomena must not be physical. It might be maintained, for example, that pain cannot be a physical phenomenon, since it is logically possible for there to be pain without matter, but logically impossible for there to be any physical phenomenon without matter.

2. The Standard Materialist Rebuttal

Against the objection that mental and physical phenomena have different essential properties, modern materialists have typically replied along roughly the following lines.

Materialism affirms that each mental state (event, process) is identical to some physical state (event, process); it affirms identity statements like "Pain = C-fiber-firings." Such identities are supposed to be *contingent* rather than necessary identities; they are supposed to be like "Water = H₂O," which is a contingent identity statement reflecting an empirical discovery.

From such a contingent identity statement it does not follow that the identified expressions have the same meaning. Water is identical to H₂O even though the terms "water" and "H₂O" have different meanings. From the identity "Water = H₂O," it does follow, of course, that water and H₂O have the same properties. However, a property that is an essential property of water *under the description* "H₂O" (like containing hydrogen) need not be an essential property of water *under the description* "water." Provided that the identity "Water = H₂O" is only contingently true (that is, that "water" and "H₂O" have different meanings), it is quite unremarkable that water and H₂O should have different essential properties (under these two different descriptions). This state of affairs guarantees that the identity "Water = H₂O" cannot be necessarily true, but it does not preclude its contingent truth.

Similarly, if "Pain = C-fiber-firings" is a contingent identity statement, then it is certainly unremarkable that pain should have some property (such as, for example, introspectability) essentially under the description "pain" but only contingently under the description "C-fiber-firings." It is part of the meaning of "pain" that pains are introspectable, but not part of the meaning of "C-fiber-firings" that C-fiber-firings are introspectable. But this no more precludes the possibility that pain is identical to C-

fiber-firings, than the fact that "water" and "H₂O" differ in meaning precludes the possibility that water is identical to H₂O. Contingent identity statements entail that the identified entities have the same properties, but (since essential properties are description-dependent) they do not entail that the identified entities have the same essential properties.²

Against the claim that mental and physical phenomena have contradictory essential properties, the typical materialist rebuttal involves insisting that a seemingly necessary statement is refutable, and is therefore actually contingent. Thus, for example, against the claim that mental phenomena are necessarily nonspatial, whereas spatial location is essential to physical phenomena, it is typically replied that we do not know a priori that mental phenomena lack definite spatial location and, consequently, that suitable experimental results could establish that thoughts, for instance, do have location in space. Lack of spatial location is thus shown (so the argument goes) to be at best a contingent property of thoughts, and the claim that thoughts have essential properties inconsistent with those of physical phenomena is thus refuted.

3. The Lockean Account of Essential Properties

The rebuttals just described depend crucially on an account of essential properties and logical necessity according to which logical necessity is always verbal necessity—that is, an account according to which necessarily true statements are just those whose truth follows from the meanings of their constituent terms. Such accounts have two important features. First, they entail that the essential properties of an entity are *relative to a description*, so that something may have a property essentially with respect to one description and contingently with respect to

another. Second, they entail that necessity and apriority coincide and thus that a statement may be shown to be contingent by showing that it is refutable. It is just these two consequences that are essential to the cogency of the standard materialist rebuttal to Cartesian criticisms of materialism.

Such accounts of necessity have been the received empiricist accounts ever since Locke, and such accounts rest upon an account of the meaning of natural kind terms that—in various versions—has been the standard empiricist account since its introduction by Locke.³ According to these empiricist accounts, the meaning of a natural kind term, or of general terms of any sort, is given by conventionally adopted criteria for telling which things fall under the term. According to such an account, for example, a term like "gold," "bachelor," "H₂O," "water," or "pain" would have *as its meaning* a set consisting of one or more properties by which gold, bachelors, H₂O, water, and pain are recognized. These properties are essential properties of gold, bachelor, H₂O, water, and pain just because they are part of the meanings of "gold," "bachelor," "H₂O," "water," "pain." Which properties should be grouped together to form the meaning of a general term is not a question of fact; it is entirely a matter of linguistic choice or decision that we associate one set of properties with each other as the meaning of a general term. ". . . our distinct species are nothing but complex ideas with distinct names attached to them"; ". . . Each abstract idea with a name to it makes a distinct species" (Locke, 1690, book III, chap. vi, secs. 8, 38).

Following Locke and Hume, the motivation for these accounts of necessity and of general terms has been antimetaphysical: the essence of a natural kind is said to consist of its "nominal essence," the conventional meaning of the term that

3 suspect

describes it, precisely to rule out of court metaphysical questions about the real essence of natural kinds. Thus, for example, Locke holds that the question whether bats are birds is a purely verbal question (1690, book III, chap. xi, sec. 7). Bats are birds if and only if the criteria conventionally adopted for applying the term "bird" apply to bats. According to such a view, it would have been literally nonsense for a seventeenth-century biologist, living in a linguistic community that considered bats to be among the paradigm cases of birds, to claim to have discovered that bats really were not birds, that they lacked the essential features shared by other birds. He could propose to change the meaning of "bird" to *make it true* that bats did not fall under that term, but there is no such thing as a fundamental principle governing the application of a general term's being mistaken: such principles are true by definition and are the basis for all necessary truths about natural kinds.

Such empiricist positions regarding general terms and necessity, particularly in their modern (and most plausible) forms, derive their plausibility from verificationist considerations. Questions regarding the real essence of natural kinds are held to lie beyond the range of possible empirical investigation and thus to involve unscientific and pointless speculation. It is impossible, according to such a view, for us to know the real essence of anything.

4. The Importance of Non-Lockean Accounts of Language and Necessity

Such accounts of necessity *de dicto*—of necessity as resting on meaning and linguistic conventions—have been accepted by the great majority of recent authors on the mind-body problem. In particular, antimaterialists who find essentialist objections to materialist theories of mind convincing typically accept the analysis of necessity upon which the rebuttals cited rest, and attack those rebuttals on other

grounds. In many respects, however, a Lockean account of necessity and of essential properties seems inadequate for a full understanding both of the issues raised by the essentialist objections and of the position of materialists themselves.

Whatever their merit may ultimately be, the essentialist intuitions in question do not seem to be fairly captured by a Lockean account of essential properties. The views that physical events are essentially spatial, or that H₂O contains hydrogen essentially, or that consciousness is an essential feature of the experience of pain, do not—at first glance anyway—seem to be judgments about meanings or linguistic usage. The philosopher to whom it seems obvious that—whatever the atomic components of water may be—these components are *essential* to water—seems, at least at first glance, to hold a position about the substance of water *itself*, not about water under some particular description ("H₂O"). Water just *is* H₂O, no matter how it is described; being H₂O is *its* essence—so the intuition goes—not merely the essence of the linguistic expression "H₂O".

If, as Kripke maintains, an account of necessity that justifies these judgments is available, then it will not merely better reflect the intuitions of antimaterialists. It will, as well, effectively disarm the standard materialist rebuttals to their essentialist objections. For, if the intuitions indicated above can be made coherent, we would have an account of necessity *de re* (not *de dicto*): an account according to which the essential properties of a thing do not depend on a particular description of it. Such an account of necessity would also preclude the strategy of showing that a statement is not necessary by showing that it represents an empirical claim that is refutable. The claim that water contains hydrogen is certainly a refutable empirical claim, but—on a *de re* account of necessity—it might be a necessary truth as well.

It is clear, therefore, that if (as, of

course, many philosophers doubt) a coherent account of necessity *de re* can be successfully defended, such an account would provide the basis for a more faithful formulation of essentialist objections to materialism than does a Lockean account. What is equally true—but less obvious—is that a Lockean account of necessity and of general terms poses difficulties for materialists as well as for their opponents. This claim may seem strange. After all, materialists typically insist that the entire body of materialist doctrines concerning mental phenomena (and other phenomena as well) are contingent empirical truths. The materialist, it seems, need not affirm that any of his doctrines are necessary truths; his interest in necessity is solely in rebutting essentialist objections to materialism and for that purpose the Lockean account of necessity is ideal.

The difficulty that the Lockean account poses arises not because materialists must defend their doctrines as necessary truths (although, as we shall see, Kripke believes that they must), but rather because the antimetaphysical philosophy of language—and, in particular, the account of natural kind terms—upon which the Lockean account rests, itself poses difficulties to the defender of materialism.

It is a crucial feature of empiricist accounts of language that questions regarding the classification of entities under general terms are always questions regarding *existing* linguistic conventions. The question whether a particular entity falls under a general term is nothing more than the question whether its properties satisfy the criteria conventionally associated with the term. What is *not* possible, according to this account, is that some entity should lack the criterial properties associated with a term by current convention, but be properly classified under that term nevertheless. It might seem that such a state of affairs is possible: the entity might possess qualities really essential to the kind referred to by the general term,

but might lack the properties by which the kind is generally identified. But this possibility is just what the Lockean account of general terms precludes. Only the nominal essence of the term is at issue in classification. We do not—because we cannot—classify things according to their own real essences or according to the real essences of the species into which they fall.⁴

In a similar way, relations of containment and identity between natural kinds turn out to be matters of current convention. The question whether two general terms name the same property, substance, or state, is merely the question whether these two terms are conventionally associated with the same criterial properties (or, perhaps with sets of criterial properties between which there is a relation of mutual meaning-entailment). It is not possible, according to a Lockean account, for two general terms with different nominal essences to refer to what is really the same property, substance, or state. If general terms referred to real essences, of course, such a situation could obtain, but since only nominal essences are involved, “each abstract idea with a name to it makes a distinct species” (Locke, 1690, book III, chap. vi, sec. 38).

The upshot of all this is that a Lockean account of necessity—and the account of general terms upon which it rests—has the effect of enshrining the *status quo* in matters of classification: it portrays the most basic standards that we employ in applying general terms as fixed by linguistic convention and immune from refutation. There is simply no such thing as discovering that our fundamental standards of classifications are wrong. We can change standards, of course (by changing the meanings of our terms). It can, perhaps, even be rational to do so—but the rationality cannot be the rationality of correcting a mistaken belief in the face of new evidence.⁵

It will now be evident why a Lockean account of general terms poses such an

acute challenge to materialism. A Lockean account of meaning enshrines our most fundamental principles of classification as definitional truths not amenable to revision. Dualism—and the principles of classification that traditional dualism supports—are among the most entrenched of our classificatory principles. Thus a Lockean materialist runs the risk of having to hold (because of his philosophy of language) that, for instance, the statement that pains are physical states of the central nervous system is not merely false but self-contradictory. After all, pains are among the paradigm cases of states that we now classify as nonphysical; if there are classificatory conventions at all (and the Lockean account insists that there are), then surely it must be a truth by convention that pains are nonphysical.

What is ruled out, it must be remembered, by a Lockean account of general terms—and by the associated empiricist epistemological outlook—is the view that, although we do not now classify pains as physical, nevertheless pain poses the same essential features as do paradigmatically physical states, and we could eventually discover that they are really physical. According to a Lockean analysis, all there is—or could be—to being physical is having the properties conventionally taken to be marks of the physical.

The view that a Lockean account of general terms, together with certain commonplace facts about current usage, is logically incompatible with materialism is not, of course, held by materialists who accept a basically Lockean account of general terms. It does not even appear to be defended—in exactly the terms presented above—by any critic of materialism. Nevertheless, closely related objections to materialism do occur in the literature, and—what is even more important to the current issue—many materialists modify their accounts of materialism to accommodate them to a Lockean account of

general terms and, in doing so, substantially weaken their own position.

Thus, for example, as we have seen, many materialists (and many of their critics) hold that materialism is committed to the truth of identity statements of the form " $M = P$ " where " M " is a general term of classification for mental states (" M " might be "pain") and " P " is some general description of a physical state, couched in obviously physical (or physiological) terms (" P " might be, for example, "The firing of C-fibers," to use a now standard imaginary example). (It will be recalled that it is just statements of this sort that play so prominent a role in the essentialist objections I am considering.)

It is conceded by materialists—in fact they typically insist on it—that " M " and " P " here have different meanings, and are associated with different criteria. The identity is supposed to be contingent, not a priori. But, according to a Lockean account of general terms, this is just the sort of identity statement that *cannot* be true. Each distinct set of criteria—each distinct "meaning"—gives rise to a *distinct* species. If " M " and " P " are general terms, different in meaning, then, it would appear, the statement " $M = P$ " is necessarily false! Now, this potential criticism of materialism is—in various guises—seriously treated in the literature. Many materialist authors are at pains to insist that contingent, non-a priori identity statements are sometimes true and known to be true. "Water = H₂O" is, indeed, the standard example.

Of course, the fact that it is possible to be a materialist, a Lockean about general terms, and a believer that water is contingently identical to H₂O does not show that it is possible to hold all these views consistently. The tension between them is revealed in the writings of recent materialists not by the admission that they are inconsistent, but rather by a special sort of exegesis offered for contingent

identity statements like "Water = H₂O," "Pain = C-fiber firings." Such identities, "theoretical identities" they are sometimes called, are not, strictly speaking, the ordinary garden-variety identities—so this sort of account goes.⁶ When, for theoretical reasons, we *identify* two terms whose ordinary rules of usage are so different, it is not strict identity we are talking about. For example, it does not follow from the theoretical identification of pain with C-fiber firings and from the fact that some pains feel vaguely cold, that some C-fiber firings feel vaguely cold.

Although proponents of this view are not always clear about the matter, their treatment of the "identity theses" really amounts to this: we do not ask, for example, "Are pains identical to C-fiber firings?" The answer to that question is "No," and the negative answer is dictated by linguistic convention. We ask instead, "Would it be reasonable, in the light of current scientific discoveries, to change our conventions so that we *can* say, 'Pain = the firing of C-fibers' without fear of self-contradiction?" This is the real issue of "theoretical identification." It is revealing that the verb for "to identify with" is so often employed in discussions of this view. We are really seen as facing the question whether or not *we* should identify pain with C-fiber firings, that is, whether or not we should adopt a new linguistic convention, to *identify* the expressions "pain" and "C-fiber firing." The issue is one of *linguistic choice*.⁸ In treating the issue this way, modern materialists continue the Lockean tradition of treating disputes over classification as "purely verbal" rather than as factual. In positivist terminology, they treat the issue whether pains are C-fiber firings as the issue whether or not to *adopt* the "meaning convention" expressed by the axiom "Pain = C-fiber firing."

Although the practice of treating ontological issues as though they were issues

of free linguistic choice, thus reformulating them as issues expressible in the "formal mode of speech," has an honorable history,⁹ the fact remains that such a gloss on materialism fundamentally distorts its claims. For better or worse, the materialist claims that mental states, events, and processes are really physical. He does not claim *merely* that we could *adopt the convention* of saying that they are. He claims they are already, anyway! If he says that pain is identical to C-fiber firings, he means it. He does not mean that *we could identify* the one *term* with the other *term*—he does not even mean that it would be rational to *adopt* such a *convention*. He certainly does not mean that, even though it is analytically false that pains are C-fiber firings, it would be convenient to change the meanings of our terms to *make* it true. What he means—for better or worse—is that pains simply *are* C-fiber firings. In his view it could, of course, be rational and (probably) convenient to say "Pain = C-fiber firings," but the rationality involved would be the rationality of accepting an important discovery in the light of new evidence, not the pragmatic rationality of *adopting* a simpler language. A "Lockean" gloss makes materialism into a mere shadow of its former self.

Worse things yet happen. As we have seen, one of those problems that face materialists is that there seem to be properties that physical states possess and mental states essentially lack, or vice versa. Thus physical states possess special locations, whereas mental states may seem to be essentially nonspatial. Similarly, some thoughts are dim, fading, or nagging,¹⁰ whereas it would seem that physical states essentially lack these properties. How is the materialist to deal with these difficulties? The "standard rebuttal" discussed in section 2 replies that, for example, it is not a necessary truth that brain states lack the property naggingness and that this can be seen by realizing that suitable experimen-

tal results (presumably those which confirm a materialist theory of mind) could show that some brain states *are* nagging, however queer that may now sound. Similarly, this rebuttal requires a defense of the claim that we could discover that, for instance, thoughts are located in the head, however queer that may sound.

As we have seen, such rebuttals depend for their cogency on a Lockean account of necessity. They proceed to establish that a statement is not necessary by demonstrating that it is not a priori. What several defenders of materialism have recognized is that this rebuttal seems to run afoul of the very Lockean account of general terms on which it rests. If there is a problem about attributing naggingness to brain states or location to thoughts, then the problem arises because such attributions are linguistically deviant enough to be counted as violating current rules of linguistic usage. If—as the Lockean account requires—there are linguistic conventions governing general terms like “has location” or “is nagging,” then linguistic normalcy and linguistic deviance must be reflections of *just those conventions*. So, there is a philosophical problem about predicating location of thoughts or naggingness of brain states if and only if such predications *violate conventions of language*, that is, if and only if the denials of such predications are, after all, really a priori, and the standard rebuttal, therefore, unsuccessful. The rebuttal works only if it is not needed.¹¹

As a substitute, many defenders of materialism have made proposals that emphasize their (perhaps unintentional) commitment to the view that the issue of materialism is (at least substantially) a question of *linguistic decision*. Thus, for example, Shaffer (1961) suggests that it would be reasonable to *change* our conventions so as to *allow* predicating location of thoughts, and Feyerabend (1963) urges that materialists recognize that they are committed to proposing such meaning

changes in order to avoid a “dualism of features.”

Where materialists ought to say that since mental events *are* physical events, they certainly *do* have locations in space, these philosophers are led by their Lockean conception of general terms to assert merely that we could *decide* to change the language to *make* materialism true. But this retreat to conventionalism is not by any means the most heroic measure taken by defenders of materialism who find themselves in this Lockean bind. Rorty (1965) concerns himself with the problem of predicating mental properties like naggingness of brain processes, which seems unavoidable if one acknowledges that some thoughts are nagging and also insists that each thought is identical to a brain process. His solution is to treat the relevant identity statements, those of the form “*M* = *P*” where “*M*” is a mental term and “*P*” a physical term, as expressing a “disappearance” form of the identity thesis. In this view, such expressions do not express ordinary identity, but rather express identity between “to put it crudely—existent entities and non-existent entities.” The statement “My thought at *t* = brain state *B*” really says that there is no such thing as my thought at *t* but that brain state *B* is what we should talk about instead. Since there are no thoughts—and hence no nagging thoughts—the problem of predicating naggingness of brain state does not arise. And, similarly, for other difficult cases of mental-physical identity.

Although ingenious, this cure may be worse than the disease. It places our materialist in the unenviable position of denying that there are thoughts, pains, feelings of joy or anguish, and so forth. It places him in an essentially untenable position. It must be pointed out that not all materialists whose philosophy of language is empiricist advocate these particular positions. Indeed, some do not even recognize the difficulties that their account of language poses to their materialism. What is

important is that it *does* pose such problems and that the available solutions (within a Lockean framework) all weaken the claims of materialists to the point that their doctrine is either untenable or not very interesting.

We have, therefore, a very interesting situation. *Both sides* in the dispute between materialists and "essentialist" dualists seem to be misrepresented if their positions are explicated in terms of a Lockean account of general terms. The materialists' position is trivialized and the essentialists' intuitions regarding necessity seem altogether misunderstood. It is for this reason that Kripke's efforts to find an alternative account of reference and necessity—and to apply it to the mind-body problem—is of such great importance for philosophy of mind.

5. "Rigid Designators" and Kripke's Account of Necessity

Kripke's discussion of materialism, with which we are primarily concerned in this essay, takes only eight of the ninety pages that constitute his development of a theory of necessity. Although it will be necessary here to provide a broad outline of the main features of Kripke's theory, I am not going to attempt to summarize all of the important aspects of its development, nor am I going to adopt a position regarding its soundness. What I do intend to show is that, assuming Kripke's account of necessity to be sound, his criticisms of materialism are not successful. I shall also indicate why the sort of account of reference that forms the foundation for his account of necessity is crucial to the defense of materialism. Although Kripke's account of necessity touches the issue of materialism primarily with respect to the issue of essential properties of natural kinds, and that of the way in which natural kind terms function in language, his exposition of necessity begins with a treatment of proper names. Since his doctrines

are clearest with respect to this issue, I shall begin there, too.

Consider the question how proper names (of people, countries, towns, and so on) refer. In certain situations, when someone uses the word "Moses," by his use of the name "Moses" he refers to the leader of the Israelite exodus. Other uses of the name "Moses" refer to other men, as when someone now says, "My neighbor, Moses, is on vacation in Vienna." Many people are, or have been, named "Moses." What is it, about, for instance, a use of the term "Moses" that *does* refer to the Israelite leader that makes it refer to him and not to someone else, or no one at all?

One influential answer (defended in various forms by both Gottlob Frege and Bertrand Russell) is closely analogous to a Lockean account of general terms. In this view, proper names like "Moses" (or any other proper name) are "disguised definite descriptions"; when someone uses a proper name, he means by the name some description or other that (if the utterance in question refers at all) holds of one and only one person. Thus, for example, if I use the name "Quine" in saying "Quine's attack on analyticity was crucial to the development of early postpositivist philosophy of science," my use of the term "Quine" refers to the philosopher at Harvard of that name because I use the name "Quine" as shorthand for some definite description of him, for example, "the philosopher who teaches at Harvard and who is the author of 'Two Dogmas of Empiricism.'" My use of the name "Quine" refers to the Quine at Harvard just because the definite description in question is true only of him. My success in communicating to a listener, in this case, depends, somehow, on a recognition of the fact that it is this (or roughly this) definite description that I have in mind (rather than, perhaps, "the man who owns the bakery on Seventh Street and Feigl Avenue"). The man who says "Quine makes good cream

puffs," and refers to Quine the baker, does so because he uses "Quine" as shorthand for just such an alternative definite description.

In the case of most proper names, it is not entirely clear just how—on this theory's account—the listeners come to understand a definite description relevantly like that intended by the speaker. In the case of historical figures like Moses, however, the solution is easier. On most "disguised definite description" accounts of such names there is a general linguistic convention that associates the term "Moses" (at least when it is used in discussing biblical history) with a definite set, or, perhaps, a "cluster" of the most important properties by which historians recognize him: male, Israelite leader, lived for a while in Egypt, led the exodus, died in Canaan, and so on. It is, according to these views, a matter of *linguistic convention* that the name "Moses" (used in the right sort of contexts) refers if and only if there is one person who has all (or, in some versions, most) of these properties. If there is such a person, "Moses" refers to him. The similarity between this sort of account and Lockean accounts of general terms is obvious. Like the Lockean account of general terms, this Russellian account of proper names has the consequence that necessity and apriority coincide with respect to certain statements involving proper names. Thus, Moses has a property essentially if and only if that property is entailed by the properties that make up the "definition" of the name "Moses." It is logically possible that Moses had brown hair (or that he had red hair) because his hair color is not mentioned in the definition of "Moses," whereas it is necessarily true (and a priori) that if there was a Moses he performed all (or, by some accounts, most) of the historical acts attributed to him in the definition. Likewise, it is logically impossible, and a priori refutable, that Moses should have

been, for instance, an officer in the Egyptian army who opposed the exodus.

Aside from any other difficulties that it may face, such an account of proper names has some provocative consequences regarding the essential properties of people. It is not surprising that Moses is not essentially brown-haired. What may seem counterintuitive is that it is both true and known a priori that there could be no possible world in which Moses sought and received a commission in the Egyptian army, became an opponent of the exodus, and died in Thebes. We have strong intuitions that such a state of affairs is logically possible, and that empirical evidence forms the only basis for our acceptance of the account that we actually believe.

Against the Russellian account of proper names—and in defense of these central intuitions—Kripke offers an alternative "causal" theory of reference for proper names. On the Russellian view, a proper name is associated with exactly the same "cluster" of defining properties in every possible world (and refers to their unique bearer, if there is one). On Kripke's view, a proper name should be seen as referring to exactly the same *individual* in every possible world in which it refers at all, whatever properties the individual may have in that world. In his view, my employment of the name "Moses" refers to Moses not because I am participating in a linguistic convention that associates with the term "Moses" some definite description, but rather because my employment of that name bears the right sort of causal connection to the historical events surrounding the giving of the name "Moses" (or whatever name it is from which the name "Moses" derives) to the man Moses. When I use the name "Moses" I, in effect, "point" back in time toward the relevant first employments of the name, and I refer to whatever was named "Moses" in these initial "dubbing" uses of the name. In this respect—according to Kripke

ke's account—my referring to Moses by the term "Moses" is more like ostensive reference—reference by pointing, for example—than it is like referring via a definite description. Similarly, when I use the name "Moses" to refer instead to Moses the literature teacher down the hall, I refer to him rather than to the historical figure not because I use his name as a disguised definite description but because my use of the name "Moses" bears the appropriate, causal relation to, for example, the events surrounding his parents' naming him "Moses," and my use of the name does not bear the right causal relation to the "dubbing" of the Israelite Moses. Certain social and linguistic conventions are involved in our use of names, no doubt. But, Kripke maintains, whatever conventionality is involved in naming, it does not result in statements about people that are true by definition.

It might seem that, according to this account, Moses, for instance, would have no nontrivial essential properties (that is, no essential properties except those that are dictated by principles of formal logic like being either living or nonliving). Indeed, such a view is compatible with the account of reference for proper names just presented. Kripke maintains, however, the plausible view that a person's parents are essential to him: that a man who, in some possible world, had different parents from those Moses had in the actual world, would not *be* Moses, whatever other properties he had. Although his account of proper names does not entail it, it does make possible such an account of description-independent (*de re*) essential properties of persons.

Here we can also see how Kripke's account of reference and necessity makes necessity and apriority distinct. Suppose that Moses' parents were Philip and Samantha. Then it would be a *necessary* truth (it would be true in all possible worlds in which Moses exists) that Moses'

parents were Philip and Samantha; and this sort of necessity obtains in the case of every person and his/her parents. But, of course, we do not know a priori who someone's parents are; it is not part of the *meaning* of the name "Moses" that Moses' parents were Philip and Samantha. Thus, "Moses' parents were Philip and Samantha," if true, is an a posteriori (indeed, a refutable) necessary truth.

Thus Kripke's account of reference and necessity (which I have examined so far only with respect to clear-cut cases of proper names) provides for the existence of a posteriori necessary truths. At least in the case of proper names, it shows how a necessary truth might be unknown to us (as when we do not know who the parents of a historical figure were) and even contrary to our most fundamental convictions (as when we are *very* sure we have correctly identified the parents of a historical figure, but we are wrong). An illustrative example of the sort of a posteriori necessary truth that plays a role in Kripke's revitalized essentialist objections to physicalism is provided by the identity statement "Cicero = Tully." Suppose that there is a man who was called "Cicero" and also called "Tully"—that each of these proper names was "given" to him. Suppose I say "Cicero is identical to Tully," and that the causal antecedents of my usages of the names "Cicero" and "Tully" are such that each of them refers to this man. Then in any possible world the referent of the terms "Cicero" and "Tully" is *that very man*, and the statement "Cicero = Tully" is, therefore, a necessary truth: it is true in every possible world in which its constituent terms refer.

For our purposes, the interest of Kripke's work arises from the extension of his account of proper names to certain other referring expressions. Kripke introduces the expression "rigid designator" for those terms that, like proper names, refer to the same thing in every possible world in

which they refer at all. Kripke suggests that various natural kind terms and terms for natural phenomena are also rigid designators, for instance, "gold," "water," "heat," "hydrogen," "pain," and so forth. Kripke denies the standard Lockean account according to which the reference of these terms is fixed by criterial attributes or defining characteristics that are "part of the meaning" of these terms. Thus, for example, a Lockean might offer an analysis of the term "heat" according to which the property by which "heat" is defined is the capacity to make us feel warm. "Heat" might be said to mean "that natural phenomenon that is present in all things that make us feel warm and that causes them to make us feel warm" or something of this sort. According to a view of this sort, the statement "Heat makes (most) people feel warm" would be both a priori and (therefore) *necessary*.

In Kripke's view, however, there is no difficulty in accepting the claim that there is a possible world in which, for instance, no animals ever developed a sensitivity to heat and in which heat produces no sensation at all in any person. Heat need not possess the analytic definition "that which warms us," nor any other analytic definition at all. As is the case with proper names, there may well have been some particular sensible characteristics by which people recognized heat when they first started calling it "heat"—just as there may have been some particular description by which Moses' parents referred to him when they named him "Moses," but these are not defining criteria for "heat" established by linguistic convention. Heat might come to lack these characteristics, or—in some possible world—might never have had them.

Of equal importance (from our point of view), the fact that "heat" possesses no analytic definition makes certain fundamental beliefs of ours about heat—which a Lockean might believe to be true by convention—refutable in principle. Thus, for

example, there would be no logical impossibility to our discovering that there were kinds of heat that, although quite intense, produced no sensation in us at all. It will be appreciated that this sort of treatment of natural kind terms represents just the sort of non-Lockean view of these terms required by the materialist who wishes to say, without self-contradiction, that certain of our most basic principles of classification associated with the terms "mental" and "physical" are (and will be shown to be) fundamentally mistaken.¹²

If rigid designators are just those terms that have the very same referent in every possible world in which they refer at all, then, it must be noticed, *some* rigid designators *may be* definite descriptions. Suppose (as is plausible) that "hydrogen" and "oxygen" are rigid designators. Consider the expression "H₂O." It is reasonable to hold that "H₂O" *means* "the compound whose molecules consist of two hydrogen atoms and one oxygen atom." Thus it is a definite description. But, since the terms "hydrogen" and "oxygen" are both rigid designators, the description "the compound whose molecules consist of two hydrogen atoms and one oxygen atom" refers to exactly the same compound in every possible world in which it refers at all. Thus, by this analysis of its meaning, "H₂O" is a rigid designator.

6. Rigid Designators and Surprisingly Necessary Identities

"Cicero = Tully," it will be recalled, is necessary if true, and this is the case for all identities of the same form involving proper names. Evidently the same is true of all identities in which the identity sign is flanked by rigid designators.

This consequence of Kripke's theory has, itself, some surprising consequences. As we have seen, philosophers who defend materialist theories of mind, and are looking for noncontroversial examples of identity statements between natural kinds, often cite the identity "Water = H₂O" as

an especially clear example. Of course, "Water = H₂O" is not an a priori truth, so, if apriority and necessity coincide (as empiricists claim), it is not necessary either. But according to Kripke's account, if, in the actual world, water really is identical to H₂O, then this identity is necessary: it holds in every possible world in which water (or H₂O) exists. Water, if it is H₂O, is H₂O *essentially*. Similarly, every chemical compound, whatever atomic constitution it has, has that constitution essentially, even if its constitution is not merely not known a priori, but not known at all. Likewise, if heat is identical to molecular vibrational kinetic energy (another standard example of an a posteriori, "contingent" identity), then heat is *essentially* molecular vibrational kinetic energy. All of these examples—if Kripke's account is sound—are cases in which a substance or natural phenomenon has an essential property independently of any linguistic convention or choice of description.

Furthermore, all of these are cases of a posteriori, *refutable* necessary truths; they provide concrete illustrations of the fact that Kripke's account of necessity—because it does not rest on a Lockean account of language—divorces apriority from necessity. In particular, they show how Kripke's account of necessity blocks any attempt to show that a statement is not necessary that proceeds by showing that it is refutable.

These features of Kripke's account—it will be recognized—make it (at least apparently) more suitable than the Lockean account for the defense of the essentialist arguments against materialism that I have been considering. Kripke's account treats essential properties of things as description-independent (thereby blocking one feature of standard materialist rebuttal) and allows for a posteriori but necessary truths (thus blocking the other). It accords with the intuition that a substance itself may have its constituents as essential features—because without just those con-

stituents it would not exist—even though the features are not specified in some analytic definition of a term referring to the substance. Thus, just as Kripke's sketch of a non-Lockean account of reference is particularly well suited to a faithful exegesis of materialist theories of mind, the corresponding account of necessity seems particularly well suited to a faithful exegesis of essentialist criticisms of those theories.

7. Essentialist Criticisms Revisited

It remains to be seen what the force of essentialist criticisms of materialism is, if they are understood according to Kripke's non-Lockean account of necessity and are, therefore, invulnerable to the standard materialist rebuttals. In order to see what the force of such criticisms is, let us first consider the case of a less controversial "theoretical" identity: "Water is H₂O."

Imagine a defender of this water/hydrogen oxide "identity thesis," who describes his position according to the standard analysis offered by materialists. He maintains that, although "water" and "H₂O" differ in meaning, the identity "Water = H₂O" is nevertheless an empirical discovery and an example of a *purely contingent* identity statement. On Kripke's account of necessity, our "identity theorist" has already adopted an untenable position. "Water" and "H₂O" are rigid designators. Therefore the identity "Water = H₂O," if it is true at all, must be a *necessary* truth. It *cannot* be contingently true. Thus the "identity theorist" must retract the claim that his identity thesis is contingent. He must live with—and, more important, he must defend—the claim that water is *essentially* H₂O, if he is to claim that water *is* H₂O at all.

This consequence of the identity "Water = H₂O" has, itself, additional consequences with which the "identity theorist" must live. The expression "H₂O" is not just a rigid designator. It is also what might be called a composition-speci-

fyng term (this terminology is not Kripke's). Whatever is H_2O must—in any possible world—be made of hydrogen and oxygen. If "Water = H_2O " is true in the actual world—and hence in all possible worlds—it then follows that having the particular molecular composition specified by " H_2O " is an *essential* feature of water. The identity theorist must, therefore, defend this claim. He must, for example, be prepared to deny the seemingly plausible claim that there is a possible world in which water does not have a molecular structure at all, because matter in that world is continuous and does not have a discrete microstructure. Similarly, the defender of the "theoretical" identity "Heat = molecular vibrational kinetic energy" must hold that there is no possible world in which heat is not, for instance, dependent on molecular motions. A possible world in which matter has no molecular microstructure is *ipso facto* a world in which *there is no heat!*

It should be emphasized that defending these claims is by no means the hopeless task that it would be if one adopted a Lockean account of necessity. Part of the point of Kripke's account is that necessary truths need not be a priori. Thus the defender of these "identity theses" does not face the hopeless task of trying, for example, to show that being made of molecules is "part of the meaning" of the word "water." The identity theorist does, however, face an important essentialist challenge: in taking the identity statements "Water = H_2O " and "Heat = molecular vibrational kinetic energy" to be contingent, rather than necessary, our "identity theorist" was not merely following current fashion. We (or, at any rate, many of those who consider such issues) have strong intuitions that, for example, water is only contingently identical to H_2O or that there could be heat in a possible world in which heat lacks a molecular microstructure. The identity theorist must provide us with

sufficiently good reasons for rejecting these strong philosophical intuitions.

The same challenge, of course, faces the materialist regarding his account of mental phenomena. If, for example, he claims that some identity statement like "Pain = C-fiber firings" is true, then he must claim that it is necessarily true. Since "C-fiber firings" is a composition-specifying term that names a kind of physical phenomenon, he must defend, for example, the claim that there could be no pains in any possible world in which there is no matter (indeed, there could only be pain in possible worlds in which matter is organized into C-fibers that fire). Similarly, he must defend the claim that it is impossible for there to be a world in which some C-fibers fire without a pain's being felt. In these cases, the identity theorist's claims run afoul of very strong philosophical intuitions indeed. As we shall see, Kripke's view is that—while the defenders of "Water = H_2O " and "Heat = molecular vibrational kinetic energy" can overcome these difficulties—mind-brain identity theorists cannot.

8. The Standard Materialist Rebuttal (New Version)

Each of the "identity theses," "Water = H_2O ," "Heat = molecular vibrational kinetic energy," "Pain = C-fiber firings," faces, at the outset, the same essentialist challenges: each asserts the identity of entities that appear to have different essential properties. The standard materialist rebuttals, which we examined earlier, depend on a Lockean account of necessity: on the doctrine that the essential properties of a thing are description-dependent, and on the doctrine that refutable statements cannot be necessarily true.

If Kripke's account of necessity and essential properties is correct, these rebuttals are ruled out. The identity theorist must hold that the identities in question represent a posteriori *necessary* truths and

that the identified entities do have (description-independently) the same essential properties. It remains then for the identity theorist to undermine philosophical criticisms of these views—to explain away the tendency to hold, for example, that water is only contingently H_2O or that heat might, in some possible world, be a nonmolecular fluid.

In order to show how this may be done, Kripke introduces what we may think of as the standard strategy for explaining away the apparent contingency of necessary a posteriori statements. The strategy involves finding a genuinely contingent statement that corresponds in the right way to the apparently contingent necessary statement, and attributing the apparent contingency of the second to the recognition of the actual contingency of the first. An example will make the strategy clear.

In the case of the apparently contingent (but necessary) identity statement "Water = H_2O ," the corresponding contingent statement might be "The cooling, tasteless, odorless, wetting liquid that quenches thirst = H_2O ." This statement is contingent, since there could be a possible world in which some other liquid than water satisfies the definite description on the left of the identity sign. The contingency of this statement, furthermore, can be adduced to explain the apparent contingency of "Water = H_2O ." The definite description "the cooling, tasteless . . ." is true of water in the actual world, and is chosen so that it describes water in terms of the properties by which it is usually recognized. If a Lockean account of terms like "water" were true, this definite description would be a candidate for the definition of "water," and would refer to water in every possible world. The intuition that "Water = H_2O " is contingent is explained as resulting from the correct judgment that the corresponding contingent sentence is contingent, together with

the mistaken belief that the definite description in it is the definition of "water" and refers to water in every possible world.

In general, in the case where the apparently contingent but necessary statement is an identity statement involving rigid designators, " $R_1 = R_2$," this strategy requires that one find referring expressions " D_1 " and " D_2 " such that, in the actual world, " D_1 " and " D_2 " are coreferential with " R_1 " and " R_2 ," but where at least one of " D_1 " and " D_2 " is a nonrigid designator that describes the properties that, in the actual world, are appropriate to the detection of the referent of the corresponding rigid designator. " $D_1 = D_2$ " will be contingent and its contingency will (together with a mistaken Lockean account of the rigid designators " R_1 " and " R_2 ") explain the apparent contingency of " $R_1 = R_2$."

In the case of apparently contingent necessary statements that are not identities, the strategy is similar. The defender of the identity statement "Heat = molecular vibrational kinetic energy" must maintain that it is a necessary truth that if there is heat, there are molecules present. The intuition that this is a merely contingent statement can be explained by adducing the corresponding genuinely contingent statement "If there is a natural phenomenon whose presence makes us feel warm, then there are molecules present." As before, the contingency of this statement, coupled with the mistaken belief that "the natural phenomenon whose presence makes us feel warm" is the definition of the term "heat" explain the apparent contingency of the necessary truth in question. In all of these cases, the apparent contingency of an a posteriori necessary statement is explained by finding a corresponding genuinely contingent statement in which one or more rigid designators are replaced by qualitative descriptions of the sort a Lockean would offer as analyses of

the meaning of the rigid designators in question. What Kripke suggests is that this standard rebuttal does not work in the case of the necessary-if-true statements to which a materialist theory of mind is committed.

9. Kripke's Argument against Materialism

According to the strategy of the standard materialist rebuttal (new version), the apparent contingency of a necessary a posteriori statement "S" is explained by finding a corresponding genuinely contingent statement "S*" in which (at least typically) one or more rigid designators occurring in "S" are replaced by definite descriptions in terms of sensible properties. In every possible world these descriptions refer (if they refer at all) to some thing or property that has in that world the same sensible properties that the rigidly designated thing or property has in the actual world. In any possible world we would be in an "appropriate qualitatively identical evidential situation" with respect to the referents of these descriptions, as we are in the actual world with respect to the referents of the rigid designators. Kripke's claim that the standard strategy fails for statements that follow from a materialist account of mental phenomena rests on the (very plausible) claim that certain mental states have as essential properties the way they feel. In any possible world, something that feels like a pain is a pain, and no pain fails to feel painful. If, in some world *W*, someone bears to an entity *e* a relation qualitatively identical to the relation we bear, in the actual world, to a pain, then *e* is a pain in *W*.

Consider, now, the materialist who defends a type-type version of the identity thesis. He holds, let us say, that each mental state is identical to a physical state of the central nervous system. In particular, he holds that pain is identical to a physical state of the central nervous system. Let us assume that the state of the central ner-

vous system in question is the firing of the mythical "C-fibers." Our materialist, then, defends the identity "Pain = C-fiber firing." If he is a typical materialist (and has not read Kripke) he will explain that this is a contingent identity statement. He will agree with the Cartesian intuition that it is logically possible that there might be C-fiber firings but no felt pain, and that it is logically possible that someone might experience a pain even though there are no C-fibers (and no other material objects, for that matter). These logical possibilities would be excluded—he will maintain—if "Pain = C-fiber firings" were a necessary identity; but it is a posteriori and hence contingent.

Of course, this position is one that Kripke's account of necessity (if it is sound, as we assume here) rules out. "Pain" and "C-fiber firings" are rigid designators of natural kinds and, therefore, "Pain = C-fiber firings" is necessary if true. The "essentialist" challenge to the materialist is to explain its apparent contingency. What he must explain, for example, are the intuitions that there is a possible world in which there are pains but no C-fiber firings and that there is a possible world in which there are C-fiber firings but no one feels any painful sensation.

This is just what Kripke says cannot be done. In the case of the identity "Water = H₂O," the corresponding problem is to explain the intuitions (a) that there is a possible world in which there is water but no H₂O, and (b) that there is a possible world in which there is H₂O but no water. The problem is soluble. To solve it, all we need to establish is that there are possible worlds *W*₁ and *W*₂ such that (a) in *W*₁, there is a liquid that is not H₂O but that has all the qualitative properties water possesses in the actual world and (b) in *W*₂, H₂O exists but fails to have the qualitative properties by which we detect water in the actual world. But, of course, it is quite reasonable to insist that such possi-

ble worlds as W_1 and W_2 exist while maintaining that "Water = H_2O " is true in every possible world.

In the case of the identity "Pain = C-fiber firings," we might expect analogous maneuvers to provide us with explanations for the intuitions (a) that there is a possible world in which there are pains but no C-fiber firings and (b) that there is a possible world in which there are C-fiber firings but no pains. By analogy to the case of "Water = H_2O ," we might expect to explain these intuitions by finding possible worlds W_1^* and W_2^* such that (a) in W_1^* there are entities that have the sensible properties that pains have in the actual world but they are not C-fiber firings, and (b) in W_2^* there are C-fiber firings but they do not have the sensible properties that pains have in the actual world.

This is the sort of explanation of the apparent contingency of "Pain = C-fiber firings" that Kripke claims is impossible. If such possible worlds as W_1^* and W_2^* exist, then the natural phenomena in W_1^* that, in W_1^* , have the sensible properties that pains have in the actual world, are not C-fiber firings, and, therefore, are not pains. The C-fiber firings in W_2^* that, in W_2^* , do not feel like pain, nevertheless are pains. But this is absurd. The sensible qualities of pains are essential to pains and definitive of them. In any possible world, anything that feels like a pain is a pain, and, thus, there is no such possible world as W_1^* . Similarly, in any possible world a natural phenomenon that is a pain must feel the way pains feel in the actual world, and thus there is no such possible world as W_2^* .

On the basis of these considerations, Kripke concludes that the standard materialist rebuttal (new version), although adequate to the defense of identities like "Water = H_2O ," must fail for those identities like "Pain = C-fiber firings" that are advanced by philosophers who defend a materialist theory of mental phenomena. Unless an entirely new sort of rebuttal can

be devised, which Kripke doubts, we must reject identities like "Pain = C-fiber firings" and the mind-brain "identity thesis" in general.

Some "identity theorists" deny that a materialist account of mental phenomena entails "type" identities like "Pain = C-fiber firings." They hold that all an identity theorist must maintain are "token" identities, which identify each particular occurrence of a mental state, event, or process with some specific physical state, event, or process. In such a view, a materialist account of mental phenomena would entail the existence of true identity statements of the form "Jones's having a pain at $T = \dots$," where the right-hand expression describes some quite specific physiological or molecular configuration. The defender of such token identity statements, Kripke observes, faces exactly the same sort of essentialist challenges as the defender of "type" identities. In either of these cases, if Kripke is right, the materialist cannot defend the required identity theses against essentialist criticisms, and materialism seems to be unworkable as an account of the nature of those mental phenomena that, like pains, seem entirely defined by their sensible properties.

10. Reply to Kripke, I:

The New Standard Rebuttal Does Work

As we shall see, the greatest weakness of Kripke's criticisms lies in the fact that—protests of its defenders notwithstanding—a materialist account of mental phenomena does not entail the sort of identity statements to which Kripke's argument applies. I shall develop this theme in section 11.

What is striking is that Kripke's arguments have an additional defect: he has underestimated the potential of the (new) standard materialist rebuttal to essentialist criticisms. Recall that, given a necessary but apparently contingent identity statement " $R_1 = R_2$," where " R_1 " and " R_2 " are rigid designators, the new standard re-

buttal requires forming a contingent identity statement " $D_1 = D_2$," where at least one of " D_1 " and " D_2 " is a description of the referent of " R_1 " and " R_2 " in terms of the symptoms typically associated with the replaced rigid designator.

The gist of Kripke's argument is that this strategy will fail if we attempt to explain the contingency of a statement of the form " $\text{Pain} = R_2$ " by finding a contingent statement of the form " $D_1 = R_2$," because if " D_1 " is an expression that designates pain in terms of just the sensible qualities that pains have in the actual world, then " D_1 " designates pain in every possible world. " D_1 " is itself a rigid designator. If we agree with Kripke (as I have for the sake of this discussion) that the sensible qualities that pains have in the actual world are essential to pain and definitive of it, then this must be right.

What seems to have been overlooked is that the successful employment of this strategy does not require that it be the less "technical" or "scientific" term in the identity that is replaced by a nonrigid designator. As we have seen earlier, we can explain the apparent contingency of " $\text{Water} = \text{H}_2\text{O}$ " by insisting on the existence of a possible world in which the corresponding contingent sentence " $\text{The cooling, tasteless} \dots \text{liquid} \dots = \text{H}_2\text{O}$ " is false. This employment of the new standard strategy relies on the fact that the sensible properties by which we typically recognize water are not essential properties of water. But neither are the standard chemical tests appropriate to the term " H_2O " logically definitive of water. It is certainly logically possible that there should be a world in which a liquid that is not H_2O satisfies all the chemical tests that—in the actual world—are reliable indicators of H_2O . Thus we could equally well have explained the apparent contingency of " $\text{Water} = \text{H}_2\text{O}$ " by appealing to the contingency of the corresponding qualitative statement " $\text{Water} = \text{the liquid that } ***$," where " $***$ " describes the standard chemical

tests appropriate (in the actual world) to the detection of H_2O .

Thus the employment of a corresponding contingent qualitative statement to explain away the apparent contingency of " $\text{Pain} = \text{C-fiber firings}$ " does not depend on the existence of a possible world in which pain does not feel like (actual world) pain or in which some nonpain feels like an actual world pain: it does not depend on a purely phenomenal description of pain not being a rigid designator. All that is required is that *either* the expression "pain" or the description "C-fiber firings" can be replaced by an appropriate purely qualitative description that does not designate rigidly.

Of course, this can be done. For any physiological or anatomical description like the imaginary "C-fiber firings" there is certainly a possible world in which something has the qualitative properties typically associated with the term in the actual world, but really is not, in this case, an instance of "C-fiber firings." Some other sort of nerve cell might, in a different possible world, look just the way C-fibers do in the actual world, or there might be specious indications that C-fibers are firing when they are really dormant. Thus we must conclude that, contrary to Kripke's suggestions, the new standard materialist rebuttal does permit one to explain the apparent contingency of mind-body identity statements like " $\text{Pain} = \text{C-fiber firings}$," and also to explain the apparent contingency of token-token identity statements like " $\text{His pain at } t = \text{such and such molecular event}$."

It is evident that the same strategy allows the explanation of the apparent possibility of worlds in which there are pains but no C-fibers, or in which C-fibers fire but no pain is felt. The apparent possibility of a world in which there are pains but no matter at all can be explained by the real possibility of a world in the following sentence is true: " $\text{There are pains but there is nothing that } +++$," where

"+++ " describes all the ways in which matter makes itself evident to the senses. A possible world in which this sentence is true would not (if pain is identical to C-fiber firings) be a world without matter, but it would be a world in which the senses functioned in such a way that nothing produced the symptoms that we typically take to indicate the presence of matter.

It would appear, then, that the (new) standard rebuttal to essentialism provides the materialist with a more powerful defense than Kripke recognizes. The materialist who insists that he is committed to the truth of identity statements like "Pain = C-fiber firings" or "Jones's pain at t = such and such molecular configuration" can successfully employ the strategy of the (new) standard rebuttal against essentialist criticisms of his doctrine. What is even more striking is that materialists are mistaken in believing that they are committed to the existence of true identity statements of either of these forms.

11. Reply to Kripke, II: Materialism without Reductionism

The materialist asserts that all natural phenomena, all events, processes, objects, and so forth, are in fact physical: all objects are composed solely of matter and all events and processes consist solely in interactions between material things. Mental events, states, and processes, in particular, differ from uncontroversially physical events, states, and processes only in the particular arrangements or configurations of matter and material forces that realize them. Pains are quite different from, for instance, earthquakes; but the difference is configurational, not constitutional. They are made of the same sorts of stuff. The strategy of essentialist objections to this claim is to insist that if true in fact, materialism must be true necessarily, and then to attempt a refutation of this latter claim.

The position that materialism must, if true, be necessarily true, rests on the

conviction that materialists are committed to the identity thesis, that is, to the truth of mind-body identity statements like "Pain = C-fiber firings," which involve rigid designators. As we have seen, the (new) standard materialist rebuttal is effective in defending these "identity theses" against essentialist criticisms. It is not, however, necessary to invoke this rebuttal. Materialism, properly understood, does not entail the sort of mind-body identity statements against which the essentialist criticisms are directed. Indeed, as we shall see, materialism poses no difficulties for most of our intuitions regarding possible relations between mental phenomena and physical phenomena. In particular, a materialist account of mental phenomena is quite compatible with the view that there are possible worlds in which mental phenomena exist but are nonphysical.

This conclusion, if sound, is significant for two reasons. First, the intuitions about necessity and possibility that underlie the essentialist criticisms of materialism are very strong ones, and the new standard materialist rebuttal does not establish that these intuitions are unfounded. It merely offers a possible explanation for them. The case for materialism is greatly strengthened if it can be shown that materialism does not even entail the sort of mind-body identity statements against which these criticisms are directed.

Second, the claim that materialism does not entail the existence of true mind-body identity statements contradicts the standard empiricist analyses of materialism. According to such analyses, materialism asserts the syntactic reducibility of the vocabulary and laws of all the sciences to the vocabulary and laws of physics. In particular, a materialist account of mental phenomena, according to such analyses, entails the definability of all mental and psychological states (or, on some accounts, all token mental and psychological states) in the vocabulary of physics.

But definitions of the sort required by such an analysis are just the sorts of identity statements linking mental and physical states that I claim materialism does *not* entail. Indeed, I shall show that the version of materialist psychology best supported by available evidence entails that mental and psychological states are *not* definable in physical terms.

The reductionist analysis of materialism shares with the empiricist account of natural kind terms discussed earlier the same verificationist and antimetaphysical motivation: it represents an attempt to "rationally reconstruct" a metaphysical question as a formal question about language. If, as I am arguing here, each of these empiricist interpretations results in a misleading account of the philosophical consequences and evidential status of materialist psychology, then we have even greater evidence for the claim made earlier that the nonverificationist treatment of natural kind terms of the sort proposed by Kripke is essential to a sound understanding of the issue of materialism, and, presumably, of other scientific issues as well.

A variety of different considerations dictate the conclusions outlined above. I shall consider them in stages.

Mind-Body Identity, Mind-Body Identity Statements, and the Apparent Necessity of Materialist Doctrines

Part of the motivation for attempts to formulate materialist psychology as an "identity thesis," aside from the empiricist reductionistic analyses, has been the desire to distinguish materialism from "epiphenomenalism," the view that mental states are not physical, but are universally correlated with distinct physical states whose causal powers explain the effects normally attributed to the corresponding mental states. Materialists, quite rightly, have been careful to insist that each mental state is identical to, not merely correlated with, some physical state.

This "identity thesis" does not, how-

ever, entail the existence of true mind-body identity statements of the sort Kripke considers, nor does it entail that materialism must be necessarily true if it is true at all. To see the distinction between these two sorts of "identity theses," consider the case of water. "Water contains hydrogen" and "Water is rare in the Gobi Desert" are both true statements, and they entail, respectively, the "identity theses" "Water is identical to a substance that contains hydrogen" and "Water is identical to a substance rare in the Gobi Desert."

Yet (assuming as I do here that Kripke's account of necessity is correct), water contains hydrogen essentially, but water is only contingently rare in the Gobi Desert. It is relatively easy to see why this is so. In the first place, neither of the "identity theses" just discussed has the form of identity statements that link rigid designators; neither is, as it stands, the sort of identity claim which must be necessary if true. Instead, each has the form "(Ex) (water = x and Px)" where "P" is ". . . is a substance containing hydrogen" or ". . . is a substance that is rare in the Gobi Desert." [Editor's note: In this anthology, the ordinary "E" is used instead of the backward "E" as the existential quantifier.]

In the first of these cases, we can conclude that water contains hydrogen necessarily only because we can find a rigid designator "R" such that (i) the identity statement "Water = R" is true in the actual world, and hence in all possible worlds, and (ii) "R" is such that anything it designates must (in any possible world) contain hydrogen. The rigid designator "H₂O" is such an "R." The corresponding situation does not obtain in the second case. Water is identical to a substance (water itself) that is, in the actual world, rare in the Gobi Desert, but (since being rare in the Gobi Desert is not an essential property of water), there is no rigid designator "R" such that (i) the identity statement "Water = R" is true in the actual world, and hence in all possible worlds,

and (ii) "R" is such that anything it designates must (in any possible world) be rare in the Gobi Desert.

We are able to show that water contains hydrogen essentially not just because water is identical to a substance that contains hydrogen, but because we are able to find another rigid designator for water (besides "water") that is formulated in a particular vocabulary (in this case, the vocabulary of chemical formulae) and that is such that whatever it names must contain hydrogen in every possible world. As we have seen, however, it is not always the case that when a statement of the form "(Ex) (x is R_1 and Px)" is true, where " R_1 " is a rigid designator, there is a second rigid designator " R_2 " such that (i) " $R_1 = R_2$ " is true and (ii) " R_2 " is such that whatever it names must have P, in every possible world.

The point is that the essentialist argument that purports to show that if mental phenomena (such as pains) are physical then they must be necessarily physical depends for its cogency on a quite specific kind of analysis of materialism. It depends on an analysis of materialism according to which, for example, the claim that pain is a physical process entails that there is a rigid designator " R_2 " that is such that (i) " $\text{Pain} = R_2$ " is true and (ii) " R_2 " is such that whatever it designates, in any possible world, must be a physical process.

As we have just seen, in the case of essential and contingent properties of water, the doctrine that pains are physical processes trivially entails that pains are identical to physical process, but this, by itself, provides no guarantee that (i) and (ii) are satisfied. It is thus perfectly consistent to affirm the "identity thesis" that pain is identical to a physical process and to deny the existence of a rigid designator " R_2 ," satisfying (i) and (ii). The philosopher who understands materialism to entail the stronger sort of "identity theses" represented by (i) and (ii) must maintain that the required rigid designators satisfy-

ing (i) and (ii) always exist. It is this claim that I deny.

*Plasticity: Compositional
and Configurational*

Let us turn now to the main issue: whether the doctrine that mental phenomena are physical phenomena entails the existence of true mind-body identity statements linking rigid designators, that is, whether it entails the truth of statements like " $\text{Pain} = \text{C-fiber firings}$." I shall concentrate first on the issue of whether a materialist account of mental phenomena entails the existence of true type-type identity statements of this kind—that is, whether materialism entails that each mental type event, state, or process is definable (by a rigidly designating expression) in a physical vocabulary. I shall turn to the issue of token-token identity statements after considering the type-type case.

My strategy is this: I shall introduce a notion of "plasticity" for type events, states, or processes, and I shall argue that the version of materialism best supported by available evidence entails that mental states admit sufficient plasticity in the way in which they are realized that it is logically possible for mental states to be nonphysically realized, even though in the actual world all mental phenomena are physically realized. By plasticity of a type of event, state, or process I understand its capacity to be realized in more than one way; the plasticity of a type of event, state, or process is indicated by the degree of variability in the particular (token) events, states, or processes that could realize it. Thus, for example, the (type) process of starting a car displays more plasticity than the (type) process of starting a 1949 Ford, because the possible token processes that could realize the first type process display greater variation (in brand of the constituent car, for example) than do the possible processes that could realize the latter.

At least roughly, we may distinguish two dimensions of plasticity (there may be more, but these are particularly relevant to the issue at hand), compositional plasticity and configurational plasticity. Compositional plasticity is displayed by a type of state, event, or process to the extent that there are possible realizations of that state, event, or process that differ in the sorts of substances or causal factors that constitute them. Configurational plasticity, in contrast, is displayed by a type of state, event, or process to the extent that its possible token realizations differ in the structural configuration or arrangement of their constituent parts, events, substances, or causal factors.

Thus, for example, the smelting of iron displays considerable configurational plasticity, since there are realizations of iron smelting involving quite different kinds and geometrical arrangements of equipment and different temporal sequences of constituent processes. Iron smelting is importantly limited, however, in its compositional plasticity: all instances of iron smelting must involve a quantity of iron. By contrast, the state of being an inscription of the English sentence "Heritability is a population-relative statistic" displays very substantial compositional plasticity: such an inscription can be written in ink on paper, carved on wood, cast in bronze, chiseled into marble, pressed into plastic, and so on. Yet the state of being an inscription of this important English sentence displays quite limited configurational plasticity: any two inscriptions of this sentence will have fundamentally similar structures; indeed, except for misspelled inscriptions, they will be isomorphic at the level of constituent letters.

An important class of states seems to possess unlimited compositional plasticity, but relatively limited configurational plasticity. "Computational states," such as being a realization of a computation of e^x for input $x = 9$, or (what is a different computational state) being a realization of

a computation of e^x for input $x = 9$ according to machine-language program P (for some definite P), or (still different) being a realization of a computation of e^x for input $x = 9$ according to a machine-language program that is a member of some definite set S of machine-language programs, all seem to possess maximal compositional plasticity: in any particular possible world, only the causal laws governing that world limit the possible composition of realizations of such computational states; such states have no essential properties that constrain the sorts of substances or causal factors that can be constituents of their realizations.

What I shall argue here—following those philosophers and psychologists who have defended the view that mental and psychological states are "functional" states of organisms—is that mental events, states, and processes are like computational states in being entirely configurational, that is, in possessing maximal compositional plasticity. It will follow that—even though mental states may always be physically realized in the actual world—there is no logical impossibility of their being nonphysically realized in some other possible world. Before turning to a defense of this claim, I must clarify some details of the notion of plasticity that are crucial to a correct assessment of the plasticity of mental states.

What is crucial to this discussion is the way in which plasticity is assessed in the case of events, states, or processes that are essentially relational. The problem can be illustrated by reconsidering the issue of the plasticity of computational states, like the state of being a calculation of e^x according to machine-language program P , for the input $x = 9$. I have said that this is a purely configurational state, and that the only properties essential to its (token) realizations are those configurational properties that are dictated by the program P . In a perfectly clear sense this is true: in any possible world W , a com-

puter capable of embodying program P , and acting on input 9 can be made of whatever arrangement of causal factors are capable—given the causal laws governing W —of realizing the required configuration of machine states and the required transitions between them.

At the same time, more is required in order for a computational state of the sort in question to be actually realized than the existence of a machine of the right sort, operating on the right sort of input, and functioning normally. It is perfectly possible for there to be one computer C_1 , which is in the state of computing e^x according to machine language program P , for input $x = 9$, and for there to be another computer C_2 , which goes through *exactly the same* succession of physical states (and, therefore, exactly the same sequence of configurational internal states) such that C_2 is *not* in the state of computing e^x according to machine language P for input $x = 9$. This state of affairs is possible because the same computer program can be used to compute quite different mathematical functions depending on the interpretation given to the language in which its inputs and outputs are represented.

The state of being a computation of e^x according to machine-language program P for input $x = 9$ has essentially a relational component. It is realized in a possible world W if and only if there is in W some arrangement of causal factors that realizes the program P and the input 9 and that is suitably related to users of its programming language whose conventions for its use are such that, with respect to them, the program P should be interpreted as computing e^x . This sort of relation to user(s) is an essential property of the computational state in question.

The existence of states, events, and processes that are essentially relational in this way forces one to refine the notion of plasticity. When one assesses the plasticity of a type of state, event, or process,

one assesses the variability in the sorts of particular (token) states, events, or processes that can realize it. The outcome of this assessment will, in the case of essentially relational states, depend on whether one adopts a narrow or broad construal of what might be termed the "scope" of the particular states, events, or processes whose variability is to be assessed.¹³

By a narrow-scope construal I understand one according to which a particular realization of a type of state, event, or process is understood to consist of those natural phenomena that actually go together to constitute the occurrence of the state, event, or process, at the time, and in the place, where it occurs, even if there are other phenomena such that if they had not occurred the particular state, event, or process would not have had whatever relational properties are essential to the type of state, event, or process in question. By a broad-scope construal, I understand one according to which a realization of a type of state, event, or process consists of the occurrence of those phenomena that constitute it according to the narrow-scope construal, together with all those occurrences by virtue of which the particular state, event, or process has the relational properties essential to the type of state, event, or process in question.

What is important to my purposes here is that plasticity can be assessed with respect to either construal, and that the results may differ according to the construal chosen. Thus, for example, the computational state I have been discussing might seem to have more configurational plasticity on the broad-scope construal than on the narrow. On the narrow-scope construal all realizations are isomorphic (actually, since a computation of a real-valued function may be nonterminating, what is really true is that any two equally long computations are isomorphic on the narrow-scope construal), whereas on the broad-scope construal there may be significant structural variety among

realizations, since there may be structurally quite different social processes that result in the adoption of the same interpretation for a programming language. More important for my purposes, the choice of broad- or narrow-scope construal will often affect assessments of compositional plasticity.

It is uncontroversial in that there are some mental states that are essentially relational in a way that precludes their being purely configurational when their scope is construed in the broad sense. For example, the state of having a vivid visual memory of the Eiffel Tower is realized by some mental processes (physical or not) only if they bear the right sort of causal relation to the Eiffel Tower (roughly, they must be caused by the subject's having in the past seen the Eiffel Tower, and the intervening causal mechanisms involved must be such that they constitute storage of visual information). Arguably, the Eiffel Tower is necessarily physical (and certainly, by Kripke's account of necessity, it is a necessary truth that the Eiffel Tower was physical when it was created). Arguably, then, every realization in the broad sense of the state of having a vivid visual memory of the Eiffel Tower must involve some physical object, and certainly there can be no realization of this mental state in any possible world in which there are and never were any physical objects.

Furthermore, as we shall see, according to certain functionalist analyses of pain a similar situation obtains. Pain (at least in the interspecific sense) is held to have as an essential feature the property of being the psychological state that is, in the species in question, typically a response to tissue damage, and typically an intermediate step in the mechanisms that lead from tissue damage to avoidance behavior. If such an account of pain is sound, and if, as seems plausible, tissues are necessarily physical, then there is an important limitation to the compositional plasticity of the state of being in pain, if

that state is given a broad-scope construal: there could be no pains in a possible world in which there has never been matter.

It will be recalled that I set out to show that, according to the best available materialist account of mental phenomena, mental states, events, and processes are entirely configurational: that is, they have no compositional properties essentially. I can now state my claim more precisely in the light of refinement in the notion of plasticity: what I shall defend is the view that, on the best available materialist account of mental phenomena, mental states are entirely configurational when they are given a narrow-scope construal and, furthermore, purely phenomenal states—states characterized solely by the quality of the experiences involved and not by their relational properties—are purely configurational on either construal of their scope.¹⁴

Let us turn now to the main task of this section. I want to show that, according to the materialist account of mental phenomena that is best supported by available evidence, mental states are entirely configurational on a narrow-scope construal and that phenomenal states are entirely configurational on any construal of their scope. I shall defend this position in a series of stages—which present evidence favoring increasingly high degrees of compositional plasticity for mental states. I shall consider the question whether materialism (in its most plausible version) entails the existence of true mind-body identity statements like "pain = . . ." where ". . ." is a rigid designator for a necessarily physical state, or, in other words, the question whether materialism entails that mental events, states, and processes are physically definable. Successive stages in my presentation correspond to the consideration of this question with respect to increasingly complex sorts of physical definitions, until, at the last stage, I conclude that the compositional plasticity of mental states rules out all possible physi-

cal definitions, however complex. The argument presented represents a rehearsal and an extension of the considerations that have led philosophers to defend "functionalist" theories of mental phenomena. For a fuller discussion of the considerations that support functionalism the reader should consult the many important recent papers defending it.¹⁵

*Compositional Plasticity, I:
Interspecific Psychological States and
Central Nervous System Plasticity*

Consider first the question whether the most plausible version of materialist psychology entails the truth of mind-body identity statements that are like "Pain = C-fiber firings" in that they identify each mental state with the operation of some quite specific neuroanatomical structure. The fact that animals of quite different species may be in the same psychological or mental state provides good reason to deny that any plausible psychology should entail such identities. It is highly unlikely that there is any quite specific neurophysiological state common to, for example, all animals that are in pain, and it is even less plausible that there is a single neuroanatomical structure whose operation is definitive of pain in all logically possible animals as well (as identities of the sort in question would entail).

Instead, what functionalists claim is essential to a mental or psychological state is not the particular physiological mechanisms that realize it, but rather the "computational" or "information-processing" role that these mechanisms play with respect to the animal's nervous system and body generally. A physiological state of an animal that plays the right sort of role in a particular animal's processing of information, and the regulation of its behavior, is a pain, on this view, even though other quite different physiological states play the same role (and, therefore, realize pain) in animals of other species. It is by no means uncontroversial just

how this functionalist position should be worked out in the case of particular mental and psychological states. It might be argued, for example, that pain has certain relational properties essentially—its causal connections to certain typical behaviors, for example; whereas, on the contrary, it might be maintained that its phenomenal qualities alone are essential to and definitive of pain (as Kripke appears to hold, at least in the case of pain in persons).

Nevertheless, it seems reasonable that the most plausible materialist response to the issue of interspecific occurrences of mental and psychological states is to adopt some sort of functionalist account: to claim that for each type of mental or psychological state, event, or process there are certain configurations of information-processing systems, or internal "programs," such that their manifestation in the body of an animal is (together with the realization of whatever essential relational properties the mental or psychological state may have) sufficient to constitute a manifestation of the mental or psychological event, state, or process in question. Functionalism of this sort entails a degree of compositional plasticity (on the narrow-scope construal) for mental and psychological states, and it rules out the claim that they possess very simple physiological definitions. Nevertheless, these considerations by themselves do not entail that mental and psychological events, states, and processes have maximal compositional plasticity. They do not rule out the possibility of a species-by-species physiological definition of mental and psychological states, nor do they rule out the possibility that mental and psychological states possess physical definitions more complex than those just considered.

The issue of species-by-species definability is important because it is quite plausible that the debate between materialists and dualists has really been con-

cerned only with mental and psychological states in man. Dualists since Descartes have often seemed to maintain the position that the attribution of mental and psychological states to animals represents a significant extension of those mental and psychological concepts that we employ in describing the states of men and have adopted, or at least been prepared to tolerate, a materialist account of mental and psychological states in nonhuman animals. An understanding of mental terms according to which they involve a certain ambiguity between their human and non-human employments is particularly appropriate to an account of Kripke's anti-materialist arguments. Kripke holds that the phenomenal qualities of pain are essential to them. Such a position is plausible if it is understood to apply to pain in man, but it is utterly implausible if it is understood to apply to pain as an interspecific mental state. Even though it is undoubtedly true that having a phenomenal quality of some sort (probably, even, having an unpleasant phenomenal quality) is essential to the interspecific state pain, it is wildly implausible that in order to be a pain, a mental state of a guppy must have the *same* phenomenal quality as a pain in man.

It is, therefore, reasonable to inquire whether the most plausible version of materialist psychology entails the existence of true identity statements each of which links a mental or psychological state in man with one quite specific physiological state: that is, whether it entails statements like "Pain in man = C-fiber firings." Once again, the considerations which support functionalism suggest that the answer is no. Identities of the sort in question would entail that it is logically impossible for any particular mental state to be realized (in man) by other than the quite specific physiological state that typically realizes it (since it is this state, presumably, with which it would be linked by the relevant mind-body identity state-

ment). There is, however, substantial evidence that such atypical realizations are not only possible but actual. For example, the most plausible accounts of certain cases of recovery from aphasia induced by brain lesions seem to be that the relevant information-processing function of the damaged tissue is taken over by parts of the nervous system that do not typically perform this function. There is no reason to doubt the logical possibility (indeed, the practical possibility in many cases) that mental and psychological states other than linguistic capacities also display a similar plasticity.

Considerations such as these make it clear that the version of materialist psychology that is most plausible in the light of available evidence does not entail the existence of true identity statements linking human mental and psychological states with quite specific neurophysiological states. They provide evidence as well for the functionalist position that mental and psychological states are closely analogous to computational states of machines. The question remains whether more complicated physiological or other physical definitions exist for mental and psychological states. Although for most philosophical and scientific purposes only finite definitions are worth considering in answering this question, for the purposes of this inquiry into the modal consequences of materialism we must consider this question with respect to infinite definitions of a sort that would have no explanatory value, and would otherwise have limited philosophical importance.

*Compositional Plasticity, II:
Realization by Mechanical Computers
and by Nonphysical Systems*

Let "Q" rigidly designate the set of all those physiological states that, in some possible world, realize pain in man. Q may well be infinite. Nevertheless we can inquire whether the most plausible version of materialism entails the identity state-

ment "Pain in man = the state of being in a state that is itself a member of Q." This statement represents the most general possible *physiological* definition of pain in man. If materialism entails the physiological definability of pain in man, and of other mental and psychological events, states, and processes, even via definitions of this complexity, then materialism does entail that mental events, states, and processes in men are necessarily physical.

Before I turn to a discussion of this issue, there is a technical question that must be resolved. In the last section I turned my attention to the question of species-by-species physiological definitions for mental and psychological states, and in particular to the question of physiological definability of such states in man. There is a possible confusion introduced when one considers the issue of the physiological definability of, say, pain "*in man*." If by "man" one intends a *biological* species, the complex physiological definition I am considering may well define pain, at least on a materialist account of mental phenomena. It is entirely plausible that creatures in some possible world, who, however much they are very like us, are nevertheless sufficiently different from us that their mental and psychological states are not physically manifested, would not be members of the same biological species as ourselves. The materialist who—setting himself up for the new essentialist challenge—affirms that even though our mental states are physical there is a possible world in which men's mental states are nonphysical is not (unless he is very careless) thereby adopting the position that such possible men would be members of the same biological species as ourselves. All he need maintain, in being faithful to our strong philosophical intuitions, is that it is logically possible that there be beings whose mental and psychological capacities are the same as ours, and whose conscious mental states are phenomenally just like ours, but

whose mental states are nonphysical. In what follows, I shall adopt the convention that in talking about mental and psychological events, states, and processes I shall be understood to be discussing events, states, and processes phenomenally just like those that occur in ordinary humans, occurring in beings whose mental and psychological capacities are those of human beings. I now turn to the issue of the physiological definability of mental events, states, and processes.

Against the definability of mental events, states, and processes in physiological terms (via even possibly infinite definitions) I argue that, according to the most plausible materialist understanding of mental phenomena, it is logically possible—indeed, even physically possible—for these phenomena to be realized by entirely inorganic mechanical computers, and, thus, that they can be realized by systems that possess no *physiological* definition whatsoever.

Two considerations indicate that machine realization of mental states is possible. In the first place, the most plausible explanation for the compositional plasticity that mental events, states, and processes seem to display in the actual world is that they can be realized in different anatomical structures because what is essential to them is their role in information processing, and their relations to other computational or information-processing structures in the same organism. Thus their compositional plasticity has the same explanation as the compositional plasticity that is displayed by what I earlier called "computational" states of computing machines. Indeed, analogies between mental states and computational states have been suggested by almost every defender of functionalism. The point is that there is no evidence to suggest that the analogy is not exact, no evidence that mental and psychological states should not be viewed as computational states of organisms. There is no reason to doubt,

therefore, that the same computational or information-processing structures could be realized in nonorganic matter as well as in animal tissue.

The second consideration concerns the available evidence for a materialist account of mental phenomena. I am inquiring here about the consequences of the version(s) of materialist psychology that are best supported by available evidence. Now, many philosophers who defend materialist theories of mind, for reasons of modesty, timidity, or methodological confusion, maintain that they defend the mere logical possibility that scientists will eventually confirm materialism, but insist that there is no currently available evidence that strongly supports a materialist account of mental phenomena. Their modesty is misplaced. It is plain that the upsurge of recent interest in materialist theories of mind reflects growing—though by no means conclusive—evidence favoring materialism. The evidence is not as “direct” as some theorists seem to require: no one has identified the particular brain mechanisms that realize a particular mental or psychological state in any higher animal. But there is substantial “indirect” evidence that favors a materialist account of mental phenomena, and materialism generally. This evidence is of three sorts: the variety of cases (brain lesions, electrode-implantation experiments, drugs with highly specific psychological effects) in which different physical and chemical changes produce different and highly specific changes in mental or psychological state; the success of modern biochemistry in the elucidation the chemistry of heredity and of other cellular processes; and the limited success of “artificial-intelligence” programs in simulating certain intellectual processes on mechanical (that is, nonorganic) computers.

The first sort of evidence is relevant because the most plausible explanation for these effects is that the various physical and chemical agents interfere with or

alter the physical realizations of quite specific mental and psychological states. The second is relevant because such cases provide evidence for materialism generally, and because they serve to refute vitalism, which is closely linked to dualism by its insistence that certain directed, purposeful, or organized structures, of the sort characteristic of living things, cannot possess a physical realization.

It is tempting to dismiss “artificial intelligence” as evidentially irrelevant. After all, only in science fiction, or in the press releases of the most boastful of its practitioners, does artificial intelligence achieve the computer simulation of the more difficult sorts of human intellectual activity. One should resist the temptation.

It must be remembered that one of the most serious objections to materialist theories of mind is the difficulty one has in even conceiving of how they might be true. We are unable to imagine exactly how an arrangement of physical parts could interact so as to manifest a feeling of pain, or so as to make a decision. Indeed, we (most of us anyway) have strong intuitions (at least some of the time) that physical realization of mental phenomena is impossible. Such “intuitions” are not to be dismissed lightly, as Kripke would be the first to insist. They should be taken seriously as *prima facie* evidence against materialism, not because “intuition” has some privileged epistemic status (it does not), nor because conceivability and possibility are the same thing (certainly they are not, at least if Kripke is right about modal logic), nor because “intuitions,” linguistic or otherwise, are the subject matter of philosophy (they are not, and philosophy is much more nearly continuous with the sciences than we ordinarily recognize). Rather, such strong intuitions should be taken seriously because what we misleadingly call “intuitions” are, quite often, instances of scientifically reasonable inductive judgments, based on observations, informed by theoretical

considerations, and amenable to revision in the light of new evidence. They are, indeed, perfectly typical examples of "theory-mediated" inductive judgments of the sort that are commonplace and essential in the proper conduct of scientific inquiry (see Boyd, 1973, forthcoming).

What this means is that the fact that a great many scientifically informed people are unable even to imagine how an ensemble of the sorts of physical systems with which they are familiar could realize a pain or a thought is itself some evidence that physical systems cannot realize such mental states. Of course, in any such case, an alternative explanation of the intuitions of impossibility is always available: failure of imagination (not necessarily culpable failure, but rather failure resulting from inadequate information or inadequate theoretical understanding). Such an explanation seems to account for the conviction shared by many eminent early twentieth-century biologists that there would not be a purely physical explanation for heredity (see, for example, Haldane, 1914).

It is, presumably, the business of materialists to offer and defend a similar explanation for the "intuition" that mental phenomena cannot be physical. Of course, the best way to do this (analogous to the discovery of DNA in the realm of cellular biology) would be actually to discover the details of the physical realization of some mental state in man. Lacking this triumph, however, it is still possible to offer less direct evidence against the reliability of these intuitions. In the first place, one can undermine the foundations upon which the intuitions of the impossibility of materialism rest. In large measure these foundations seem to consist in the conviction that certain traits characteristic of mental and psychological states—rationality, self-directedness, purposefulness, ingenuity, self-organization, adaptability, and the like—cannot be realized by a purely "mechanical" system. Against

this conviction, materialists must claim that it rests on an unduly narrow conception of the range of possible mechanical systems. Evidence for this materialist rebuttal is provided by every case in which it is established that physical systems can realize some traits of this sort, which would previously have been thought to lie in the realm of the necessarily nonphysical. This consideration explains the central importance of advances in biochemical explanation of cellular phenomena as evidence for a materialist account of mental phenomena: the organized, self-reproducing, and adaptive cellular processes that now have been chemically explained are precisely the sorts of processes about which antimaterialist intuitions would have led (indeed, did lead) philosophers and scientists to doubt their physical realizability. Advances in "artificial intelligence" have also made a crucial contribution along the same lines to the defense of materialism. Although there has certainly not yet been machine realization of anything like human consciousness or problem-solving capacity (small wonder, given the puny size of modern computers when compared with the brains of even non-human mammals), it is still true that "artificial-intelligence" programs have realized the simulation of many problem-solving capacities of just the sort that would have been (were!) thought to lie in the exclusive realm of the nonphysical. It would be a mistake to underestimate the effect these achievements have had, *and ought to have had*, in undermining the *prima facie* force of antimaterialist intuitions.

The foundation of antimaterialist intuitions can be undermined from a different direction: one can advance theoretical understanding so that what was inconceivable becomes conceivable. We are still not at the stage where we can imagine exactly how human intelligence, for instance, or painful feelings, could be physically realized. Nevertheless, our capacity to imagine that they could be so realized

has been considerably increased by the success, even the partial success, of cognitive psychologists in offering "information-processing" accounts of cognitive functions. Such accounts (even those of psychologists who are extremely doubtful about the fruitfulness of actual computer "modeling" of human cognitive functions) rely heavily on analogies between human cognition and "information processing" by machines, as the prevalence of computer-derived terminology ("feedback," "memory-limitation," "parallel processing," "subroutine," "information retrieval," and so on) in the working vocabularies of cognitive psychologists shows. Whenever a mental or psychological phenomenon is explained by a theory that rests on an analogy to the operation of physical mechanisms, as information-processing theories do, it becomes more reasonable to attribute the seeming impossibility of physically realized mental life to a lack of sufficient theoretical understanding.

Both of these rebuttals to antimaterialist intuitions, it will be noted, turn on the assumption that human mental and cognitive processes are "computational" or "information-processing" processes of the sort that can be, in principle, realized by nonorganic systems. Thus not only does the positive evidence of actual plasticity of the human nervous system indicate that mental events, states, and processes are machine realizable, but the most plausible materialist rebuttals of antimaterialist intuitions also rest on this assumption. I conclude, therefore, that the most plausible materialist psychology entails the machine-realizability of mental events, states, and processes, and thus that it entails that such events, states, and processes *do not* possess physiological definitions, however complex.

One important point must be made about the conclusion drawn here that mental and psychological states are "computational" states or organisms. There is

no single notion of "computational state"; as we have seen, such states may be thought of as characterized merely by the function they compute—by input-output relations—or they may be thought of as characterized by one or more aspects of the particular "program" or arrangement of causal factors that realizes the computation. Most emphatically, it is with respect to the latter understanding that I mean to defend the view that mental events, states, and processes are computational. In the literature cited earlier, there have been objections raised to various versions of functionalism on the grounds that they did not acknowledge the fact that certain mental states, such as pain, have their qualitative features as essential properties. This criticism is certainly warranted if functionalism is understood to assert that mental states are characterized—as computational states—by the "function" they compute, that is, by just the role they play in linking sensory stimulation to behavior. What the materialist must maintain is that, for each sort of mental or psychological event, state, or process, there exists a definite class of possible (temporarily extended) patterns of interaction, or "programs," such that the realization of a member of this class by a physical system is a necessary and sufficient condition for the physical realization of all the nonrelational properties essential to the mental state, event, or process in question. According to this view, for example, there are certain configurations such that whenever they are realized by a physical system, whatever substances compose it, the qualitative feeling of pain is manifested. This is a very bold claim, and the evidence for it is not entirely conclusive. It is, nevertheless, dictated by the version of materialist psychology that is best supported by available evidence and it is probably one of the important grains of truth in various formulations of functionalism.

We will now inquire whether the most

plausible version of materialism entails that mental and psychological events, states, and processes in man possess physical definitions of any sort, physiological or not. As before, we may formulate the most general possible physical definition for any mental or psychological state. In the case of pain, for example, let *W* be the set of all those physical states that, in some possible world, realize pain. We will inquire whether the most plausible version of materialism entails the identity "Pain in man = the state that is realized by all and only members of *W*," and similar identities for other mental and psychological states in man.

By way of an answer, we are now in a position to see that materialism (in its most plausible version) entails that mental states are purely configurational, on a narrow-scope construal—that is, that their nonrelational essential features place no logical limitations whatsoever on the sorts of causal factors that may realize them. This conclusion is virtually dictated by what I have said so far. If all that is required to realize the nonrelational essential properties of any particular mental state is the physical realization of a certain configuration by any sort of matter whatsoever, and if this is true because mental states are computational or information-processing states, then there is no good reason for supposing that the same mental state would not be realized if the same configuration were realized by nonphysical causal factors. I conclude, therefore, that the most plausible materialist psychology entails that mental and psychological events, states, and processes are purely configurational on their narrow-scope construal, and that purely phenomenal events, states, and processes are, therefore, purely configurational on a broad-scope construal as well.

Possibilities, Possibilities . . .

We are now in a position to offer a new reply to certain essentialist chal-

lenges to materialism. The essentialist critic challenges the materialist to explain the strong philosophical intuitions that we have that materialism is only contingently true, and that there are possible worlds in which mental phenomena are not physical. The reply I now offer is that this claim is compatible with the most plausible version of materialist psychology, which itself entails just the right sort of compositional plasticity for mental events, states, and processes. It is, indeed, fully compatible with a plausible materialist psychology that there should be a possible world in which there is no matter at all, but in which there are events, states, and processes that have all the nonrelational properties essential to the mental events, states, and processes manifested in the actual world.

There is one essentialist challenge that I must consider here (before turning to the issues raised by token-token identity statements). Kripke, in moving against the materialist who asserts an identity statement like "Pain = C-fiber firings" maintains not only that we have strong intuitions that there could be pains without C-fiber firings, but also that we have strong intuitions to the effect that there could be C-fiber firings (or any other physical events) but no pains. He challenges the materialist to explain away this intuition as well. Of course, this particular form of the essentialist's objection does not apply to the most plausible versions of materialism, which do not entail the truth of such mind-body identity statements. But a closely related version of the same objection can still be formulated. The plausible materialist maintains that there are certain types of physical event (C-fiber firings, for instance) that have all the nonrelational properties essential to pains (in Kripke's view, these are all the essential properties of pain). In particular, the materialist maintains that it is impossible to have C-fibers fire without a pain being felt. But the same intuitions to

which Kripke appeals in criticizing the identity statement "Pain = C-fiber firings" operates in this new case as well: we have strong intuitions that—contrary to the dictates of the most plausible materialism—it is possible to have C-fibers fire without a pain being felt.

As I have noted earlier, this challenge can be met by the standard materialist rebuttal (new version) whose strength Kripke has underestimated. In this case, however, one cannot reply, as I did in the case of the strong intuition that pains are possible in a world without matter, by accepting the intuitions as sound. What can be done, however, is to show that this particular application of the new standard materialist rebuttal is entirely unproblematical. To see this, consider a simpler case of the physical realizations of functional states, ordinary electronic computers. Suppose that one considers the circuits of a simple computer that, in fact, computes the square of whatever input is entered. Unless one happens to have previously studied circuits just like the one in question, one will have the strong intuition that the mess of wires could be turned on, the input 9 entered, the whole thing operate normally, and the result 15 be displayed subsequently. One cannot tell by looking (at least not easily) just what function a particular circuit will compute, or, even, that it is the circuit capable of realizing a computation at all. It is difficult to deduce the function computed by a computing device just from a specification of its internal structure (indeed, if one considers the generalized question of computers with unlimited memory, there is no generally effective procedure for recognizing computers that compute the square of their inputs). One has the intuition that the circuit in question could operate normally and still compute something other than the function that—in fact—it must compute, because it is quite easy to visualize the circuit while visualizing successive states of the display that are in-

compatible with the computations it in fact performs. Only if the functional capacity of a circuit leaped right out at us when we reflected on its physical appearance would we not have these intuitions. Yet it is unproblematically logically impossible for a circuit of the sort I am considering to operate normally (for an input of sufficiently small size) and not compute the square of its input. We have mistaken intuitions in cases of this sort because it is possible to visualize the structure of a computer without becoming aware of the function it computes, so that we can visualize the structure of the computer, think of it as operating normally, and visualize a behavior of its input-output systems that it cannot possibly realize, without any sense of contradiction. There is nothing odd or problematical about such an analysis of mistaken modal intuitions.

Of course, according to a functionalist account of the sort I am defending here, mental states like pain are computational and are subject to the same potential mistakes regarding their essential features. We can, indeed, visualize any anatomical configuration we like, and think of it as functioning normally, without recognizing in it the realization of any mental state whatsoever. But this is just what we should expect if mental states are computational states, and there is no reason to doubt that the (new) standard materialist rebuttal fully explains our mistaken intuitions in this case.

Thus I conclude that, insofar as intuitions concerning the essential properties of types of mental events, states, and processes are concerned, the materialist who adopts a functionalist position can—because he is not committed to any mind-body identity statements—accept the central Cartesian claim that it is logically possible for there to be mental life without matter, and—because his analysis of mental states is functional—he can offer an extremely plausible explanation for the seeming possibility that the physical sys-

tems that in fact realize mental states could exist without these mental states being realized. I turn now to the issue of the essential properties of token mental events, states, and processes.

*Token States and Quite Specific
Molecular Configurations*

Kripke raises another class of essentialist criticisms against the materialist. The materialist must, he suggests, hold that each individual ("token") mental event, state, or process is identical to some quite specific physiological or molecular event. Thus, for example, it might be true to say "Jones's pain at t = the firing of fibers $F_1, F_2 \dots$ " Yet, for any particular pain, for example, and any particular specific set of molecular motions, we have the intuition that each could occur without the other. Kripke suggests that what I have been calling the standard materialist rebuttal (new version) is inadequate to explain these intuitions, which are, of course, incompatible with the token-token identity statement in question. What should the functionalist materialist have to say regarding these intuitions?

In the first place, of course, we have already seen that the (new) standard rebuttal is sufficient to explain these intuitions. Furthermore, the compositional plasticity of type mental states that the functionalist analysis entails allows us to agree with the intuitions to the extent of saying that a pain phenomenally just like Jones's pain at t could indeed have been realized even though fibers F_1, F_2 , and so forth, had not fired; indeed, a pain phenomenally just like Jones's pain at t could have been realized in a world with no matter at all. Apparently, we must deny the possibility that those very fibers could have fired at t without realizing Jones's pain at t , but the discussion in the previous section shows that, in this case, application of the (new) standard rebuttal is profoundly unproblematical.

Although the responses just indicated

rebut the antimaterialist intuitions in a fashion perfectly adequate to the defense of materialism, there are plausible arguments that seem to support an even more satisfactory resolution of the challenge they represent. In the first place, it is a mistake to understand materialism as entailing that each token mental event, state, or process is identical to some quite specific molecular or physiological event, state, or process. The compositional plasticity that types of mental events, states, and processes display is mirrored in a corresponding transworld compositional plasticity for token mental events, states, and processes. Furthermore, it is plausible that this compositional plasticity is sufficient to make it logically possible that a token mental event, state, or process that is physically realized in the actual world could be nonphysically realized in some alternative possible world.

Some philosophers seem to have understood transworld identification of physical events, states, and processes to require microscopically identical molecular realizations, or something very close to it. They may reason as follows: an actual world physical event is nothing over and above the motions of the molecules that constitute it (ignoring the issue of essential relational properties, at least). Thus it is identical to those motions, and in any possible world in which it occurs it must be nothing over and above those very motions. The same conclusion can be reached as a consequence of the seemingly innocuous doctrine that if events in two possible worlds are identical they must have *exactly* the same causes. After all, each molecular motion that is part of the physical realization of an event is one of its causes, albeit perhaps a very minor one.

A number of examples show that this view is mistaken, and that it is, therefore, mistaken to affirm that physical events are typically identical to those smaller constituent motions (such as molecular

motions) that constitute them. In the first place, such a transworld criterion of identity is strikingly at variance with the actual world criteria of identity for physical objects. A car remains the same car in the actual world if its generator (a constituent part) is replaced, and it is hard to see why the same plasticity should not obtain across possible worlds. The man who says, "Jim replaced the generator in his car yesterday, but he might not have replaced it at all," certainly seems to be—correctly—describing a possible but not actual state of the very car that Jim owns in the actual world. Similarly when he says (perhaps in response to the question whether the new generator caused Jim's actual world accident), "Jim's accident could have occurred even if he had not replaced the generator," he certainly seems to be describing a possible world in which the very same accident occurs but in which no replacement of the generator has occurred previously. But this will be a possible world in which the molecular constituents of the accident are different, since these constituents certainly include the motions of all the engine parts including the generator.

Finally, historical events certainly seem to admit a corresponding plasticity. The historian who says, "World War II would have ended earlier had the Allied powers not adopted the 'Unconditional surrender' slogan," certainly seems to be talking about a possible outcome for the very same war which, in the actual world, ended in August 1945. It would be absurd to insist that the materialist should resist this conclusion on the ground that World War II, like all other events, is entirely physical and therefore must have *exactly the same* physical realization in each possible world. Instead, the materialist should maintain that many kinds of physical events are like physical objects in displaying transworld plasticity.

Nevertheless, there are some sorts of physical events that do not admit this sort

of plasticity. Let m be a particular molecule, and t be a particular trajectory (understood as a function from historical times to spatial locations). The token event m 's moving along t is an example of a physical event that has exactly the same physical realization in every possible world in which it occurs. Let us be sure that token mental events, states, and processes do admit the plasticity typical of physical events, states, and processes.

Let P be an actual world person with a normal lifespan. Consider a possible world W in which clever Martians slowly but systematically replace parts of P 's nervous system with nonorganic structures that bear to the remaining parts of his nervous system the same functional relations the replaced parts bore (they produce, for example, the same electrical potentials, and the same chemical substances), without interrupting in any way P 's phenomenal states, or affecting in any way his behavior or the way in which his (increasingly nonorganic) nervous system processes information. Eventually, we can imagine, this process is completed and P now has a nervous system that is entirely artificial and we can imagine that he lives out his life with no change whatsoever in his phenomenal life, behavior, or information-processing states. The compositional plasticity already established for type mental states ensures that such a state of affairs is logically possible (even though it may be technically or physically impossible). For every token mental event, state, or process of P in the actual world, there is a corresponding token mental event, state, or process of P in W that has the same phenomenal features and is connected in exactly the same way with P 's behavior, his other mental states, and plays the same role in P 's information processing. Are these corresponding mental events, states, and processes identical?

It seems extremely plausible to say that they are. We (or at any rate, I) have strong intuitions that what is essential to

the transworld identification of token mental events, states, and processes are the roles they play in the whole history of the subject's phenomenal experience, behavior, and cognitive processes. Token events (or states, or processes) that, in two different possible worlds, play exactly the same role, in this sense, in the mental life of the same person are identical. The difference in their physical realization is as irrelevant here as the difference in generators is irrelevant to the transworld identification of cars. The compositional plasticity of types of mental events, states, and processes produces transworld compositional plasticity in token events, states, and processes.

If this position is sound, as it seems to be, then the functionalist materialist can accept the intuition that any particular actual world mental event, state, or process could exist in a possible world in which the physical events that manifest it in the actual world do not occur. Indeed, there seems to be no barrier to the functionalist materialist's asserting that any particular actual world mental event, state, or process could be—in some other possible world—nonphysically realized. All one need do is to invoke a possible world in which the systematic replacement of parts of the central nervous system involves their replacement by non-physical causal factors with the capacity to influence the other parts of the central nervous system in a way that exactly simulates the function of the replaced part (which we can imagine becomes deactivated). Finally, the same considerations appear to admit the possibility that certain kinds of actual world token mental events, states, or processes might be realized in some other possible world even if the body of the subject no longer exists.

All of these latter considerations are speculative: perhaps they push the notion of a possible world to the breaking point. Their soundness is not essential to a defense of materialism, but the very fact that

they can be plausibly defended shows the extent to which functionalist versions of materialism avoid the sort of essentialist criticisms that Kripke offers.

Identity and Composition

The conclusions (not merely the most speculative ones) of the last section depend on a claim that merits further consideration. Let p be a particular pain and let c be the particular molecular process that realizes it in the actual world. Then, although it certainly makes sense to say that p is nothing over and above c , it is a mistake to claim that p and c are identical. Furthermore, although functional states, events, and processes seem to provide especially clear examples of this phenomenon, it is generally true that particular physical events, states, and processes, and physical things, of course, are not identical to the molecular arrangements that realize them.

Although this claim is widely accepted with respect to physical things, many philosophers find it much less plausible with respect to events, states, and processes. In the case of mental events, states, and processes, their conviction is partly explained by the vocabulary that has come to be used in formulating key issues in the mind-body problem. Part of the evidence for materialism consists in observed correlations between symptoms of various physical and chemical changes in the body and symptoms of corresponding changes in mental states. Indeed, some philosophers have thought that convincing evidence for materialism would rest on the establishment of correlations between symptoms of quite particular mental states and symptoms of the physical states that realize them. Regarding this sort of evidence, the question has been raised whether this correlation of symptoms, if it were observed, could not be explained in a fashion compatible with dualism by attributing the correlation of

symptoms to a universal and lawlike correlation between nonphysical mental states and corresponding physical states of the central nervous system. The issue between this interpretation of the data and the materialist interpretation has come to be described as the issue of whether the corresponding mental and physical states are identical or (as the dualist suggests) merely correlated.

As we have seen, this way of putting the question is fundamentally misleading. The issue is not identity versus correlation, but composition versus correlation. The issue is whether the physical state associated with a mental state constitutes or realizes the mental state in question, or whether, on the contrary, it merely correlates with it. The tendency to put the issue in terms of identity rather than composition may rest to some extent on an ambiguity of the English verb "to be." Suppose that Jones's pain at t were realized by the firing of fiber f . In a perfectly good sense, it would be correct to say "Jones's pain at t was just the fiber f 's firing at t ." This is correct in just the same sense that it is correct to say "In the early days, Fort Dingbat was just a circular pile of stone and rubble; only under the administration of Colonel Graft did it grow into the imposing edifice we see today." In neither case is the "was" the past tense of the "is" of identity; if it were, then Fort Dingbat could not have been added to and remained the same fort, nor could Jones's pain at t have been realized in some other structure than the fiber f . Neither are these "was"s instances of the past tense of the "is" of predication. Instead, they represent what might be called the "is" of composition (or of realization or constitution). There is nothing dubious about this use of the verb "to be," but it has nothing significant to do with identity.

An additional reason for rejecting the identity of a token event, state, or process with its actual realization is provided by the observation that the same set of molecular motions may realize several differ-

ent token events, states, or processes, as well as several different types of events, states, or processes. For example, suppose that the set of molecular motions that realizes Jim's pain at t constitute the firing of a particular C-fiber f . Then these motions realize at least three different token states: Jim's pain at t , the firing at t of f , and the token event that satisfies the description "the motion of m_1 along t_1 , and the motion of m_2 along t_2 , and the motion . . ." where the enumeration describes the precise trajectory for each of the molecules involved in the actual world of Jim's pain at t . These token states are not identical (as one can easily see by reflecting on the fact that, for any pair of them, there is a possible world in which only one is manifested), so they could hardly all be identical to the particular set of molecular motions in question (the third token state is, of course, identical to just that set of motions).

These considerations have the effect of making token events, states, and processes seem less like stereotypical "individuals" and more like type events, states, or processes—more like "universals"—in that a token event, for example, may have more than one instance (although in different possible worlds), to none of which it need be identical. This would be worrisome were it not for the fact that consideration of the issue of reidentification of individuals (physical things, for example, and people) over time shows that individuals are not very much like the convenient philosophical stereotype of individuals either.

Some philosophers (for instance, Geach, 1957) have suggested shifting whatever mystery there is in these facts about token events, states, and processes (and objects, as well) from the realm of metaphysics (physics?) to the realm of language by maintaining that some or all identity statements are incomplete unless they involve a sortal that specifies the sort of sameness relation in question. Thus, a particular set of molecular motions might

be the *same mental state* as Jones's pain at *t*, the *same physiological state* as the firing of fiber *f* at *t*, the *same chemical state* as something else, but it would be a misuse of language to say that it was the same state (*simpliciter*) as any of these since the expression "the same state as" is incomplete and without definite meaning unless qualified by some such adjective as "mental," "physiological," or "chemical." There is, in other words, no such thing as identity *simpliciter* between states.

Such a solution has, in my view, little merit since it leaves to be explained what the various qualified sameness relations have in common, and the only sound answer to that seems to be that the sentence "*a* is the same *F* as *b*" is true just in case the *F* that *a* realizes is identical (*simpliciter*) to the *F* that *b* realizes. At any rate, whatever the merits of this linguistic maneuver, it does illustrate a tendency that provides the historical basis (and much of the current plausibility) for the view that materialism entails mind-body identity statements of the sorts I have been discussing. I have in mind the tendency to formulate or "rationally reconstruct" "metaphysical" statements as non-"metaphysical" statements regarding syntactic features of appropriate terms and sentences. It is this tendency that underlies both the "Lockean" conceptions of language and necessity and the view that the claims of materialist psychology are to be analyzed as claims about the syntactic reducibility of the terms and laws of psychology to the terms and laws of physics. The results of this paper should provide additional evidence—assuming such evidence is needed—of the bankruptcy of such "antimetaphysical" positivist positions in the philosophy of language and the philosophy of science.

12. Conclusion

I remarked earlier that Kripke's account of natural kind terms not only served to clarify essentialist criticisms of materialism, but also provided the founda-

tions of an account of language that is crucial to the defense of materialism. The rival Lockean account of general terms, I argued, had the effect of treating as unrefutable linguistic conventions whatever principles of classification into natural kinds are most fundamental to current practice. Such an account makes any theoretical claims that involve fundamental change in classification (of the sort materialist psychology proposes) false by definition! Thus the Lockean account of natural kind terms appears to be incompatible with (or at least to pose very serious difficulties for) the view that materialist psychology is even logically possible, much less confirmable. Understanding natural kind terms as referring ostensively, as Kripke proposed, makes it possible to hold that some or all such terms are used to refer to kinds whose essential properties are to be discovered by scientific (or other) investigation, but not by reflection on linguistic conventions. Such an account of natural kind terms seems essential to a satisfactory account of the many cases in which research on a particular natural kind may turn up facts that are at variance with the most fundamental earlier beliefs regarding that kind, and such an account is therefore essential to a satisfactory understanding of the claims of materialist psychology. The implications for the philosophy of science of this ostensive view of reference of general terms are only now being investigated (see, for example, Goldstein, 1977; Putnam, 1975b), and much work in this area has yet to be done.

It must be remarked that in holding, with Kripke, that there is an ostensive aspect to the way in which the reference of natural kind terms and other scientific terms is fixed one need not necessarily hold, as Kripke seems to, that there is *no* descriptive component to the reference-fixing "apparatus" of such terms, nor need one hold that the scientifically relevant notion of essential property has sufficiently clear application outside the actual

world to support the account of logical necessity offered in Kripke's work. Nevertheless, such an account of logical necessity does seem to be necessary in order to capture the force of essentialist criticisms of materialism, and some sort of ostensive account of the reference of general terms seems essential to any account of the possibility of genuinely novel scientific discoveries.

Thus, if Kripke's criticisms of materialism fail, they nevertheless provide us with the opportunity to examine the strongest versions of the sort of essentialist criticisms they represent, and the account of language on which they rest will undoubtedly play an important role in the development of postpositivist philosophy of science.

Notes

1. An especially clear and compelling expression of this optimism regarding the eventual physical explicability of mental phenomena is given by Smart (1970).

2. For argumentation of this sort, see Place, 1970, and Smart, 1970.

3. See Locke, 1690, book III, especially chap. iv. I do not mean to suggest that Locke, among the traditional empiricists, has had the most direct influence on the philosophy of language of contemporary logical empiricism. That honor certainly falls to Hume. Locke has priority, nevertheless, and his especial concern for the issue of essential properties makes it only fair to cite him in a discussion of Kripke's views. See Kretzmann, 1968, for an interesting discussion of Locke's semantic theory.

4. See Locke, book III, chap. vi, especially sections 8, 9, 10. This discussion does not deviate importantly from contemporary empiricist accounts.

* ~~~~~ 5. The reader will recognize the similarity between this empiricist doctrine regarding change of classificatory principles and T. S. Kuhn's treatment of change of "paradigm." Despite his intention to be antiempiricist, Kuhn's relativistic treatment of paradigm change depends on just the sort of empiricist conception of language and conceptual change that I am discussing here (see Kuhn, 1962).

6. The term "theoretical identity" appears in Putnam, 1960. The best explicated version of this doctrine seems to be that defended by Nagel (1965). His "postscript" (Nagel, 1971) repudiates this particular view, but Nagel, 1965, remains the clearest exposition of it. Cornman (1962) speaks of a kind of "cross-category identity" immune from some applications of Leibniz's Law. The doctrine that the "identity thesis" is not really an *identity* thesis is also implied by various treatments of the issue of predicating mental predicates of brain states and physical predicates of mental states; see Cornman, 1962; Feyerabend, 1963; Rorty, 1965; Shaffer, 1961.

7. Rorty (1965) does insist that "language changes as empirical discoveries are made," but he does not discuss in detail the relation between this fact and empiricist theories meaning. Similarly, Putnam (1967) talks about "not wholly unmotivated extension of ordinary language" as underlying theoretical identifications.

8. For an explicit version of this view, see Feyerabend, 1963.

9. See, for example, Carnap, 1937, 1956; Schlick, 1959.

10. These examples are from Cornman, 1962.

11. Many philosophers propose to cope with this difficulty, and related difficulties regarding the issue of discovery versus "meaning change," by adopting a modification of the Lockean account of general terms according to which the meaning (and the reference) of a general term is fixed by a cluster (often a "law-cluster") of criteria a sufficient number, *but not all*, of which must remain unchanged if meaning is to be preserved. These "clusters" consist of the most deeply entrenched of the criteria actually employed in the typical usage of the term.

It is by no means clear that this strategy succeeds any better than those discussed later in this section. In the first place, it has proven remarkably difficult to spell out just which subsets of criteria in a "cluster" are sufficient for meaning-preservation (indeed, it is hard to say just what goes into the "cluster" and what does not). The issue seems so sensitive to conflicting intuitions and judgments that one wonders if any doctrine along these lines is available except that we decide by convention *after the fact* what changes in criteria we *will*

take to have preserved meaning. This solution would hardly help the beleaguered Lockean materialist.

Finally, even if a satisfactory account of sufficiency were available, it is by no means clear that materialism could escape the charge that it involves unacceptable changes in the "clusters" associated with both physical and mental terms. Consistent materialism would—as everyone recognizes—require us to say a very large number of things that now seem so strange that many would consider them senseless.

12. Strictly speaking, we may distinguish two features of Kripke's account of natural kind terms, *both* of which are required for an adequate defense of the revisability of fundamental principles of classification. In the first place, of course, a non-Lockean causal theory of reference for natural kind terms is a prerequisite for any account according to which natural kinds are not defined by conventionally fixed criteria of classification. A causal account of the mechanism of reference for natural kind terms, of the sort Kripke offers, *does not*, however, by itself preclude the view that natural kinds are defined by conventionally fixed (and logically necessary) criteria of classification. It is perfectly consistent to maintain—following Locke—that natural kinds are defined in just that way while adopting a causal theory of reference for the terms that refer to them. According to such a modified Lockean account, the first users of a natural kind term *T* would establish (by arbitrary convention) a set of logically necessary and sufficient defining properties for the kind referred to by *T*. A subsequent use of the term *T* would refer to the same natural (artificial?) kind if and only if it bore the right sort of causal relation to the original "dubbing" use of *T*.

Such an account accepts Locke's understanding of what a natural kind is (roughly, the extension of a conventionally fixed set of criteria) but a Kripkean account of the way in which terms refer to these kinds. This account is essentially Lockean. Indeed, it is probably the most plausible version of the Lockean account since it makes it easy to explain how someone can use a natural kind term to refer to a conventionally defined natural kind even though he does not himself know what the conventional definition is. "I especially want to see the gnus. I haven't the foggiest idea what

they are like." It remains, in particular, an account according to which the essence of a natural kind is its "nominal essence": according to which logical necessity is always verbal necessity.

The second component, then, in Kripke's account of natural kind terms is the claim that they refer to real rather than to nominal essences. Kripke's achievement is to show how such an account can be integrated into a plausible theory of reference. He does not, however, offer a fully developed non-Lockean account of natural kinds themselves. Such accounts will be required before the relevance of Kripke's work to the philosophy of science can be fully appreciated.

13. I am grateful to William Wimsatt and Sydney Shoemaker for helpful discussions regarding this point. I use the term "scope" at Wimsatt's suggestion.

14. It is, of course, arguable that there are no purely phenomenal states: that certain relations to bodily behavior are essential for every sort of mental state. It is beyond the scope of this paper to explore that issue. I here accept, for the sake of argument, Kripke's view that their qualitative character is essential to, and definitive of, for example, pains in man.

15. For discussions (pro and con) of functionalism see Armstrong, 1968; Block, 1978; Block and Fodor, 1972; Fodor, 1965, 1968; Putnam, 1975a, 1975c, 1975d, 1975e; Shoemaker, 1975. Block (1978) makes the interesting observation that, although functionalism entails that mental states cannot be identified with particular states of the central nervous system (since functionalism entails that mental states could be realized by nonbiological states), many authors nevertheless take functionalism to *support* the view that mental states are identical to physical states of the central nervous system. The discussion that follows, together with the earlier parts of section 11, can be taken as a resolution of the puzzle that Block raises. What materialists should claim is that mental states are *in fact* central-nervous-system states but that their having a central nervous system realization is not essential to them. Such an account is exactly like the one defended here: that mental states are identical to contingently physical states.

References

- Armstrong, D. M. 1968. *A Materialist Theory of Mind*. London: Routledge and Kegan Paul.
- Block, N. 1978. "Troubles with Functionalism." In C. W. Savage, ed., *Perception and Cognition. Issues in the Foundations of Psychology. Minnesota Studies in the Philosophy of Science*, vol. 9. Minneapolis: University of Minnesota Press.
- Block, N., and J. A. Fodor. 1972. "What Psychological States Are Not." *Philosophical Review* 81, no. 2 (April):159-181.
- Boyd, R. 1973. "Realism, Underdetermination and a Causal Theory of Evidence." *Nous* 7, no. 1 (March):1-12.
- . Forthcoming. *Realism and Scientific Epistemology*. London: Cambridge University Press.
- Carnap, R. 1937. *The Logical Syntax of Language*. London: Routledge and Kegan Paul.
- . 1956. "Empiricism, Semantics and Ontology." In *Meaning and Necessity*. Chicago: University of Chicago Press.
- Castañeda, H.-N. 1967. *Intentionality, Minds, and Perception*. Detroit: Wayne State University Press.
- Cornman, J. 1962. "The Identity of Mind and Body." *Journal of Philosophy* 59 (August 30):486-492.
- Feyerabend, P. 1963. "Mental Events and the Brain." *Journal of Philosophy* 60, no. 11: 295-296.
- Fodor, J. A. 1965. "Explanation in Psychology." In M. Black, ed., *Philosophy in America*. Ithaca: Cornell University Press.
- . 1968. *Psychological Explanation*. New York: Random House.
- Geach, P. 1957. *Reference and Generality*. Ithaca: Cornell University Press.
- Goldstein, A. 1977. "Meaning Reference and Theory-Change." Ph.D. dissertation, University of Michigan.
- Haldane, J. S. 1914. *Mechanism, Life and Personality*. New York: Dutton.
- Kretzmann, N. 1968. "The Main Thesis of Locke's Semantic Theory." *Philosophical Review* 77:175-196.
- Kripke, S. 1972. "Naming and Necessity." In D. Davidson, ed., *Semantics of Natural Language*. New York: Humanities Press.
- Kuhn, T. S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Locke, J. 1690. *An Essay Concerning Human Understanding*. New York: Dover, 1959.
- Nagel, T. 1965. "Physicalism." *Philosophical Review* 74, no. 3:339-356.
- . 1971. "Physicalism," with "Postscript." In D. Rosenthal, ed., *Materialism and the Mind-Body Problem*. Englewood Cliffs, N.J.: Prentice-Hall.
- Place, U. T. 1970. "Is Consciousness a Brain Process?" In C. V. Borst, ed., *The Mind/Brain Identity Theory*. London: Macmillan.
- Putnam, H. 1960. "Minds and Machines." In S. Hook, ed., *Dimensions of Mind*. New York: New York University Press.
- . 1975a. "Brains and Behavior." In *Mind, Language, and Reality. Philosophical Papers*, vol. 2. London: Cambridge University Press.
- . 1975b. "The Meaning of Meaning." In *Mind, Language, and Reality. Philosophical Papers*, vol. 2. London: Cambridge University Press.
- . 1975c. "The Mental Life of Some Machines." In *Mind, Language, and Reality. Philosophical Papers*, vol. 2. London: Cambridge University Press.
- . 1975d. "The Nature of Mental States." In *Mind, Language, and Reality. Philosophical Papers*, vol. 2. London: Cambridge University Press.
- . 1975e. "Philosophy and Our Mental Life." In *Mind, Language, and Reality. Philosophical Papers*, vol. 2. London: Cambridge University Press.
- Rorty, R. 1965. "Mind-Body Identity, Privacy and Categories." *Review of Metaphysics* 19 (September):24-54.
- Schlick, M. 1959. "Positivism and Realism." In A. J. Ayer, ed., *Logical Positivism*. Glencoe, Ill.: Free Press.
- Shaffer, J. 1961. "Could Mental States Be Brain Processes?" *Journal of Philosophy* 58 (December 21):813-822.
- Shoemaker, S. 1975. "Functionalism and Qualia." *Philosophical Studies* 27 (May):291-315.
- Smart, J. J. C. 1970. "Sensations and Brain Processes." In C. V. Borst, ed., *The Mind/Brain Identity Theory*. London: Macmillan.