

MENTAL PROPERTIES*

THE identity thesis, let us agree, is the doctrine that every mental property (e.g., being in pain) is identical to a first-order physical property (e.g., having firing C-fibers). Prevailing opinion seems to be that the identity thesis is false. The most popular arguments for this opinion are the multiple-realizability argument and the Nagel-Jackson knowledge argument. I shall try to show that these arguments are unsatisfactory as they stand and that their problems are incurable. For better arguments, opponents of the identity thesis must look to the history of philosophy, specifically to two closely related arguments traceable to the writings of René Descartes—the modal argument and the certainty argument. Although these arguments are also unsatisfactory as they stand, I shall try to show that (unlike their more fashionable cousins) they can be successfully reformulated. But for a refutation of the identity thesis in its full generality, they will need to be woven together into a single new argument.

A unifying theme of the paper is that each of the four arguments is plagued by doubts that originate in scientific essentialism (SE), the doctrine that there are a posteriori necessities (e.g., water = H₂O). If SE suitably generalizes beyond physical property identities, each of the four arguments is at best inconclusive. True, Saul Kripke has tried to save the modal argument from this kind of aggressive SE, but I shall argue that his defense (and much of the discussion of it) is misconceived.

I. THE MODAL AND MULTIPLE-REALIZABILITY ARGUMENTS

The weakest version of the modal argument is this:

It is possible that someone could be in pain and not have firing C-fibers.¹

If the property of being of feeling pain = the property of having

* I presented this paper at the National Humanities Center, UCLA, University of Western Washington, Davidson College, University of California at Irvine, USC, and the Eastern Division of the American Philosophical Association. My warmest thanks to Rogers Albritton, Erik Anderson, Louise Antony, John Bickle, Brad Cohen, Kit Fine, John Heil, Mark Hinchliff, Paul Hovda, Frank Jackson, Norman Kretzmann, Stephen Leeds, Samuel Levey, Janet Levin, Michael Levin, Joseph Levine, Brian Loar, William Lycan, Ruth Marcus, Jennifer Murphy, Terence Parsons, Michael Peirce, C. D. C. Reeve, William Reinhardt, Georges Rey, Sydney Shoemaker, Gary Stahl, Carol Voeller, and Peter Woodruff.

¹ Possibility in this sense does not imply causal or nomological possibility. 'Possible' is used here for the much weaker modality according to which something is possible if and only if it is necessary or contingent.

firing C-fibers, then it is not possible that someone could feel pain and not have firing C-fibers.

Therefore, the property of being in pain \neq the property of having firing C-fibers.

The argument is valid, and the second premise is undeniable. So the soundness of the argument comes down to the first premise. This premise is intuitive, so why not accept it straight off? SE is the main source of doubt.² After all, we could give a formally analogous argument for the conclusion that the property of being water \neq the property of being H₂O. The first premise of this argument would be: it is possible that there could be a sample of water that is not a sample of H₂O. At least initially, this seems intuitive. But there are powerful SE arguments against this premise (and in turn, the conclusion). What is or is not possible regarding water cannot be decided a priori by our modal intuitions; empirical science is required. The worry is that perhaps SE can be generalized from physical properties, such as the property of being water, to mental properties, such as the property of being in pain. Until this question of the scope of SE is settled, the modal argument is in doubt.

I said that this version of the modal argument is weak. To see why, notice that it does not refute the identity thesis in its full generality. After all, the conclusion of the argument is consistent with the thesis that having firing C-fibers is a sufficient condition for being in pain. So identity theorists are free to maintain that being in pain is identical to the disjunctive property formed from the property of having firing C-fibers and the other (perhaps infinitely or indefinitely many) first-order physical properties that are sufficient for pain. Alternatively, identity theorists are free to maintain that being in pain is identical to some first-order physical property that is common to the first-order physical properties that individually are sufficient for pain. To rule out such theses, the first premise of the modal argument must be strengthened. For example, the new premise might be: it is possible for there to be someone who feels pain but does not have firing C-fibers *and* someone who has firing C-fibers but does not feel pain. (This challenges the thesis that having firing C-fibers is a sufficient condition for being in pain.) Alternatively, the new premise might be something like: it is possible for there to be a being that feels pain but has no body. (This premise challenges the thesis that the envisaged disjunctive property is a

² A radical response is to deny that intuitions have any evidential weight. For a rebuttal, see my "The Incoherence of Empiricism," *The Aristotelian Society*, Supp. Vol. LXVI (1992): 99-138.

necessary condition for being in pain.) But, as with the weak modal premise, SE calls these stronger modal premises into question.

The multiple-realizability argument is also designed to refute the identity thesis in its full generality. One common formulation of it resembles the weak modal argument in that its first premise is a generalization on the latter's modal premise. The new premise is that there are infinitely (or indefinitely) many possible physical "realizations" of pain: C-fibers, silicon fibers, et cetera. Because this is a modal premise, however, SE raises the same doubt about it as about the first premise in the weak modal argument.

This formulation has a further problem. As with the first premise of the weak modal argument, this new premise does not, by itself, prevent the property of being in pain from being identical to the disjunctive property formed from the infinitely (or indefinitely) many possible "realizations" of pain. Because the aim of the multiple-realizability argument is to refute the identity thesis in its full generality, there is need for special auxiliary premises (which are not needed for any of the modal arguments). But these auxiliary premises turn out to be problematic in their own right. For example, one familiar version³ uses an auxiliary premise to the effect that, if there is no finite first-order physical definition of a given property, the property is not a first-order physical property. (For alternative auxiliary premises, see (i)–(iii) in the next paragraph.) But there are counterexamples. Suppose that the primitive general term '*H*' was introduced with the following reference-fixing remark: "*H* is a property with a rigid extension, and something has *H* if and only if it is a space-time point occupied by some subatomic particle." Intuitively, *H* is not a second-order physical property; it certainly has no noncircular finite second-order physical definition. But it would be very odd to say that *H* is not a physical property. So we seem forced to the view that *H* is a first-order physical property, but one that has an infinitary definition: $Hu \text{ iff}_{\text{def}} u = x_1y_1z_1t_1 \text{ or } u = x_2y_2z_2t_2 \text{ or } \dots$. Hence, a counterexample. (If one replaces 'finite' with 'explicitly stateable', 'humanly recognizable', etc., there are even more counterexamples.) The larger point is that unstateability by finite beings does not, in general, establish ontological conclusions.

There are also nonmodal formulations of the multiple-realizability argument. The main premise is that a mental property like being in

³ E.g., Jerry A. Fodor, "Explanations in Psychology," in *Philosophy in America*, M. Black, ed. (Ithaca: Cornell, 1965), pp. 161–79, esp. p. 174; Hilary Putnam, "The Mental Life of Some Machines," *Mind, Language, and Reality: Philosophical Papers, Volume 2* (New York: Cambridge, 1975), pp. 408–28; see also Putnam, "Philosophy and Our Mental Life," *ibid.*, pp. 291–303, esp. p. 299. Note that several distinct versions of the argument are alive in contemporary discussions. Some people even seem to equate it with the weak modal argument.

pain actually has very diverse physical "realizations." The auxiliary premises are (something like) the following: (i) if a property has very diverse physical "realizations," any first-order physical definition of it must be disjunctive; (ii) "natural" (versus "Cambridge") properties do not have disjunctive definitions; and (iii) mental properties like being in pain are "natural" properties.⁴ There are many problems with this argument. First, its main premise is subject to a genuine (versus merely skeptical) problem of other minds as it applies to creatures in species physiologically very unlike ours (for example, fish). Are we sure that what they feel is really pain, sure enough to base the rejection of the identity thesis on it? Naturally, advocates of the nonmodal formulation have their answers ready, but the debate threatens to end in a stalemate. At this point, the most convincing—and least vulnerable—move would be to observe that, surely, it is at least possible that physiologically very dissimilar creatures could feel pain. But this is a modal premise, which takes us back to a modal formulation. A second drawback is that the auxiliary premises invoke a metaphysical framework ("natural" versus "Cambridge" properties) which seems to many people (not me) to be as controversial as the identity thesis itself. Another problem is that (i) is open to counterexamples: a property like being 100 degrees Celsius has very diverse "realizations"; yet it has a nondisjunctive definition.⁵ Premise (ii) is also open to doubt. Definition-by-cases is a standard method of definition in logic, mathematics, philosophy, and science, and such definitions are standardly written out in the form of disjunctions. It is implausible to claim that every property requiring that kind of definition is a "Cambridge" property. A final problem is that some identity theorists could invoke a sophisticated SE point to deny (iii). Specifically, they could draw a parallel between 'pain' as it applies to physiologically diverse species and 'jade' as it applies to chemically diverse compounds.⁶ If there is no shared "natural"

⁴ E.g., Putnam, *ibid.*, and "The Nature of Mental States," *op. cit.*, pp. 429–40; Ned Block and Fodor, "What Psychological States Are Not," in *Readings in Philosophy of Psychology, Volume I*, Block, ed. (Cambridge: Harvard, 1980), pp. 237–50; Fodor, "Special Sciences, or the Disunity of Science as a Working Hypothesis," in Block, pp. 120–33. These formulations rely implicitly on (something like) the "natural"/"Cambridge" distinction. Putnam and Fodor also seem to accept (something like) the following auxiliary premise: if $\lceil \dots F \dots \rceil$ qualifies as a good explanation for a certain purpose (e.g., understanding human motivation) and $\lceil \dots G \text{ or } H \text{ or } I \dots \rceil$ does not, then the property of being $F \neq$ the property of being G or H or I . This premise is based on a fallacy akin to the paradox of analysis. See below.

⁵ See, e.g., Jaegwon Kim, "Physicalism and the Multiple Realizability of Mental States," in Block, pp. 234–6.

⁶ See, e.g., Kim, "Multiple Realization and the Metaphysics of Reduction," *Philosophy and Phenomenological Research*, LII, 1 (March 1992): 1–26.

property in the latter case, how can we be sure that there is in the former? Again, stalemate threatens. The most convincing—and least vulnerable—response is to retreat to the very weak modal premise that, surely, it is at least possible that some common “qualitative feel” could be shared by creatures in physiologically diverse species. But as soon as modal premises are invoked, the SE worry reemerges.

The conclusion is that the multiple-realizability argument requires auxiliary premises that are quite problematic; in addition, it runs into the SE worry. The modal argument—both weak and strong—also runs into that worry. But at least it is free of problematic auxiliary premises.

II. THE CERTAINTY AND KNOWLEDGE ARGUMENTS

The certainty argument is roughly this: I am certain that I am in pain, but I am not certain that I have firing C-fibers; therefore, pain and firing C-fibers are not identical. A mirror image of the certainty argument is the knowledge argument, which goes roughly as follows: suppose I know all of a being's physical properties; it does not follow that I would thereby be in a position to know its mental properties; therefore, the physical properties and the mental properties are different. The knowledge argument is thus “bottom/up” whereas the certainty argument is “top/down.”

According to traditional logical theory, the above version of the certainty argument is ambiguous, having both an intensional (“opaque”) reading and an extensional (“transparent”) reading. On the intensional reading, the argument is invalid. (Analogy: I am certain that I have pain; I am not certain that I have the experience under discussion in lecture three of *Naming and Necessity*; therefore, pain \neq the experience under discussion in lecture three of *Naming and Necessity*. On the intensional reading of this argument, the premises are true but the conclusion false.) Alternatively, on its extensional reading, the above version of the certainty argument is equivalent to: pain is such that I am certain that I have it; firing C-fibers are such that I am not certain that I have them; therefore, pain \neq firing C-fibers. This argument is valid, but to know the second premise, I would already have to know that pain \neq firing C-fibers, the very thing the argument is supposed to show. So on its extensional reading the argument is question-begging.⁷

⁷ In “Reduction, Qualia, and the Direct Introspection of Brain States,” this JOURNAL, LXXXII, 1 (January 1985): 8–28, Paul Churchland makes analogous criticisms of the Nagel-Jackson argument. But his criticism collapses when it comes to the intensional-abstraction formulation of that argument.

To save the certainty argument from these difficulties, let us try shifting to an intensional-abstraction formulation. Consider the following modal argument form: (i) $\Box Fx$; (ii) $\neg \Box Gx$; therefore, being $F \neq$ being G . Arguments with this form are valid despite the fact that 'F' and 'G' occur intensionally. Now consider the intensional-abstraction formulation of the certainty argument:

- (1) I am absolutely certain that I am in pain.
 - (2) I am not absolutely certain that I have firing C-fibers.
-
- (3) Therefore, being in pain \neq having firing C-fibers.

In view of the above, the mere fact that 'am in pain' and 'have firing C-fibers' occur intensionally does not explain what is wrong with this argument. Mere intensionality is not the problem. To see the real problem, contrapose (2) and (3) (i.e., interchange the (2) and (3) and change their signs). The new argument is formally equivalent to the original. The problem is that this new argument has the same form as the standard propositional-attitude version of the *paradox of analysis*. For example,

I am absolutely certain that computable functions are computable functions.

Being a computable function = being a recursive function.

Therefore, I am absolutely certain that computable functions are recursive functions.

This argument is invalid. The first premise is true, as is the second: being a computable function is the same thing as being a recursive function. (If not, what is it? Substitute your favorite definition. Or is computability an undefinable primitive? If so, choose a definable property.) But the conclusion is false: I am *not* absolutely certain that all computable functions are recursive. Because the argument is invalid and because the intensional-abstraction formulation of the certainty argument is formally equivalent, it is invalid as well. This is the real problem.

Let us look more closely at the propositional-attitude version of the paradox of analysis. The puzzle it raises is this: if being a computable function is just being a recursive function, then since I am certain that computable functions are computable functions, how could I fail to be certain that computable functions are recursive? The standard informal explanation is this: I am not certain of how to define what it is to be a computable function. That is, I am not certain that a function is computable iff_{def} it is recursive. Of course,

it is difficult to devise an intensional logic that adequately captures these phenomena. But this does not belie the phenomena.

The intensional-abstraction formulation of the certainty argument is formally equivalent to the propositional-attitude version of the paradox of analysis and is therefore invalid. Identity theorists who accept SE should diagnose the situation thus. Being in pain is in fact having firing C-fibers. Yet I can be certain that I am in pain and not certain that I have firing C-fibers. The reason is that I do not know the relevant definition: x is in pain iff_{def} x has firing C-fibers. This definition is a *scientific definition* (akin to the scientific definition of water: x is water iff_{def} x is H₂O) whereas we previously considered a mathematical definition. But this difference has no bearing on the argument's form and resulting invalidity. Indeed, matters are worse, for as opposed to the mathematical case, it is impossible for me to bridge the gap a priori. The reason is that scientific definitions are essentially a posteriori.

The same criticism can be made of the knowledge argument: it is formally equivalent to a propositional-attitude version of the paradox of analysis and, hence, is invalid. Identity theorists may agree with Jackson (Nagel *mutatis mutandis*) that there are things that Mary does not know about color experience before her release.⁸ Their explanation can be that Mary does not know the relevant scientific definition. If she did, she would have all the information needed to know everything there is to know about experiencing red. It is impossible for her to know this definition, however, without having experienced red and having used such experience evidentially; indeed, the need for evidential use of such experience is (part of) what makes the definition essentially a posteriori. Because her ignorance of the definition is entirely consistent with the thesis that experiencing red is identical to a physical property, the knowledge argument is no threat to the identity thesis.

With respect to knowing what it is like to experience red, identity theorists might prefer a slightly more complex story. Suppose upon her release Mary exclaims, "So *this* is what it is like to experience red." Four items are open to scientific definition: the relation holding between u and v such that u experiences v ; the relation holding between x and y such that x is what y is like; red; and *this*. From

⁸ Mary is a neurophysiologist confined to a black-and-white environment; she knows every physical fact relevant to experiencing red, but she never experiences red until her release. Frank Jackson, "What Mary Didn't Know," this JOURNAL, LXXXIII, 5 (May 1986): 291-5. Note the structural similarity between the knowledge argument and G. E. Moore's "open-question argument" (associated with the naturalistic fallacy).

these scientific definitions—plus physical facts relating the right-hand sides of these definitions to one another—one can deduce the new fact which Mary discovers, namely, that *this* is what it is like to experience red.

Are these scientific definitions (*u* experiences *v* iff_{def} . . . ; etc.) *physical facts*? The question is terminological. Identity theorists are always free to answer that they are not physical facts but rather *definitional facts*—namely, facts of scientific definition. In turn, they may hold that, although before her release Mary knows every physical fact about experiencing red, she does not know every fact about experiencing red: specifically, she does not know these definitional facts and various physical-cum-definitional facts, which are logical consequences of the physical facts plus these definitions. On this view, all the facts are either physical, definitional, or physical-cum-definitional. This is all that identity theorists ever wanted to claim.⁹

So do mental properties have the envisaged scientific definitions? That depends on the scope of SE.

III. KRIPKE'S DEFENSE OF THE MODAL ARGUMENT

I now turn to this question of the scope of SE. The arguments supporting SE rely on intuitions; without them SE would be unjustified. Consider the famous twin-earth intuition: if all and only samples of water here on earth are composed predominantly of H₂O and if, traveling to another planet, we were to find samples of a stuff that is macroscopically like water but composed of XYZ (\neq H₂O), those samples would, intuitively, not be water. Suppose that this and kindred intuitions are correct, and suppose that all and only samples of water are as described. Then, we may conclude that in all actual and counterfactual situations something would be composed of water if and only if it were composed predominantly of H₂O. In turn, we may conclude that, necessarily, water = H₂O.

⁹ Suppose Mary is told the scientific definitions before her release. She would not fully understand them—not the way people with color experience do. But this does not show that there are two sets of definitional facts; there is only one. This can be seen as follows: if Mary is asked to tell us the definitional facts after her release, she would give exactly the same answers she would give if she were asked before her release. Nor would her words have changed their meanings; she would just understand them more fully. With this in mind, identity theorists should put their point thus: anyone who knows with understanding all the physical and definitional facts is in a position to know with understanding all the facts.

Functionalists may use these points to rebut the knowledge argument: if Mary knew and understood the relevant functional definitions (i.e., of *this*, experiencing, being like, red), she would be in a position to know all the facts (including the functional-cum-definitional fact that *this* is what it is like to experience red); but it is impossible for her to know and understand these things without first having the relevant experiences.

But there is a problem. Before the advent of SE, we had a host of anti-SE intuitions, for example, the intuition that it could have turned out that some samples of water contained no hydrogen. What are we to make of the conflict between pro- and anti-SE intuitions?

Proponents of SE have two responses. First, they could simply declare that anti-SE intuitions are mistaken whereas their own pro-SE intuitions are correct. But critics of SE could simply meet this response by stating that things are the other way around. The result would be a stalemate. To avoid it, proponents of SE must turn to the second response. According to it, widespread conflict among our intuitions is only an appearance. All, or most, of our intuitions are correct. Despite their correctness, however, many are misreported. When we try carefully to rephrase our (apparently) anti-SE intuitions to make them consistent with our pro-SE intuitions, we succeed. But when we try to rephrase the latter to make them consistent with the former, we fail. Accordingly, the stalemate is broken in favor of SE.

Kripke and his followers have used two rephrasal strategies to defend SE. The first turns on an alleged pragmatic equivocation in the kind of possibilities at issue. When we report our pro-SE intuitions (for example, twin-earth intuitions), what we say is strictly and literally true, and we are reporting ordinary possibilities. But when we report our apparently anti-SE intuitions, we confuse ordinary possibility with the possibility of a certain kind of epistemic situation. For example, when we say 'It could have turned out that some samples of water contained no hydrogen', what we say is strictly and literally false. The intuition is true but incorrectly reported. The correct report would be something like this: 'It is possible for there to be a population of speakers in an epistemic situation qualitatively identical to ours and they use the expression "water" to refer to something other than water and/or they use the term "hydrogen" to refer to something other than hydrogen'. As Kripke¹⁰ remarks in connection with the Hesperus/Phosphorus case:

Now this seems very strange because in advance, we are inclined to say, the answer to the question whether Hesperus is Phosphorus might have turned out either way.

¹⁰ *Naming and Necessity* (Cambridge: Harvard, 1980). Incidentally, Kripke tells us, "[I]t could have turned out that *P* entails that *P* could have been the case" (pp. 141–2). For ease of exposition, I shall slide between 'It could have turned out that *P*', 'That *P* could have been the case', and 'Possibly *P*'. No question will be begged, for I could always revert to the first form.

And so it's true that given the evidence that someone has antecedent to his empirical investigation, he can be placed in a sense in exactly the same situation, that is a qualitatively identical epistemic situation [to ours], and call two heavenly bodies 'Hesperus' and 'Phosphorus', without their being identical. So in that sense we can say that it might have turned out either way (*ibid.*, pp. 103–4).

Generalizing from these examples, we arrive at the following schema for applying this rephrasal strategy: 'It could have turned out that A ' is to be rephrased as 'It is possible that a population of speakers in an epistemic situation qualitatively identical to ours would make a true statement by asserting ' A ' with normal literal intent'. Consider our intuition that it could have turned out that there were samples of water containing no hydrogen. The rephrasal comes out true because in the envisaged population of speakers 'water' might not name water but rather XYZ or 'hydrogen' might not name hydrogen but rather X. When rephrased thus, the original apparently anti-SE intuition is plainly consistent with the thesis that, necessarily, water = H₂O.

For the second rephrasal strategy, it is best simply to quote Kripke:

In the case of some necessary *a posteriori* truths, however, we can say that under appropriate qualitatively identical evidential situations, an appropriate corresponding qualitative statement might have been false. The loose and inaccurate statement that gold might have turned out to be a compound should be replaced (roughly) by the statement that it is logically possible that there should have been a compound with all the properties originally known to hold of gold. The inaccurate statement that Hesperus might have turned out not to be Phosphorus should be replaced by the true contingency mentioned earlier in these lectures: two distinct bodies might have occupied, in the morning and the evening, respectively, the very positions actually occupied by Hesperus-Phosphorus-Venus.

I have not given any general paradigm for the appropriate corresponding qualitative contingent statement. Since we are concerned with how things might have turned out otherwise, our general paradigm is to redescribe both the prior evidence and the statement qualitatively and claim that they are only contingently related. In the case of identities, using two rigid designators, such as the Hesperus-Phosphorus case above, there is a simpler paradigm which is often usable to at least approximate the same effect (*ibid.*, pp. 142–3).

Kripke's "simpler paradigm" goes as follows. Suppose that ' R_1 ' and ' R_2 ' are co-designating rigid designators whose designatum might have been fixed by the nonrigid (i.e., contingent) designators

' D_1 ' and ' D_2 ', respectively. When we report an apparently anti-SE intuition with 'It might have turned out that $R_1 \neq R_2$ ', our intuition is correct, but it is to be rephrased as 'It is possible that $D_1 \neq D_2$ '. On its standard narrow-scope reading, the latter sentence is consistent with the SE thesis that, necessarily, $R_1 = R_2$. For ' D_1 ' and ' D_2 ' are only contingently co-designating. For example, 'It might have turned out that water \neq H₂O' might be rephrased as: 'It is possible that the clear thirst-quenching stuff \neq the such-and-such chemical compound'. The latter is consistent with the thesis that, necessarily, water = H₂O, for there is a possible situation in which there could be a unique clear thirst-quenching stuff that is not a such-and-such chemical compound.

Although the two rephrasal strategies bear a superficial resemblance to one another, they are clearly not identical. For example, as indicated, there is at hand a rather precise rule for applying the first strategy; Kripke indicates that there is not at hand a precise rule for applying the second strategy.

This brings us to Kripke's modal argument against the identity thesis (*ibid.*, pp. 144–55). It is just a traditional modal argument together with an auxiliary argument to the effect that the second rephrasal strategy fails to deflate the modal intuition upon which the argument rests. Three points are in order. (1) The argument is plainly inadequate as it stands, for Kripke neglects one of his own rephrasal strategies, namely, the first one. (2) Kripke holds that the second strategy fails to deflate the indicated modal intuition, for any description ' D_1 ' that might serve to fix the reference of 'pain' would, like 'pain' itself, rigidly designate pain (*ibid.*, pp. 151–3). Specifically, every application of the second rephrasal strategy to the sentence 'It is possible that pain \neq firing C-fibers' leads to a sentence (e.g., 'It is possible that the feeling that feels like this \neq firing C-fibers') that strictly and literally entails the original sentence. So the intuition retains its original force. There are two errors in this reasoning. The first is in not recognizing that, depending on the circumstances, nonrigid designators ' D_1 ' could be used—and surely sometimes are used—to pick out pain. For example, 'the feeling to which I am most averse' and 'the feeling for which I have greatest spontaneous hate' pick out pain, but they do so only contingently. After all, they might have picked out nausea.¹¹ The second

¹¹ In "Kripke's Argument against the Identity Thesis," this JOURNAL, LXXII, 6 (March 1975): 149–67, Michael Levin tries to make this point with the topic-neutral description 'what is going on in me when my skin is being damaged'. But this description does not apply uniquely to pain.

error, noted by Richard Boyd,¹² arises even if one overlooks the first error. For it would still not follow that every application of the second rephrasal strategy to 'It is possible that pain \neq firing C-fibers' yields a rephrasal that entails the original sentence. To obtain a rephrasal that does not, one merely needs a nonrigid description ' D_2 ' that contingently fixes the reference, not of ' R_1 ', but of ' R_2 '; that is, one merely needs a nonrigid description ' D_2 ' that contingently fixes the reference, not of 'pain', but of 'firing C-fibers'. And they are easy to find. Identity theorists thus have a way of honoring the original intuition used in the modal argument without threatening their thesis that, necessarily, pain = firing C-fibers. By fixating on ' R_1 ' (i.e., 'pain'), Kripke neglected to consider ' R_2 ' (i.e., 'firing C-fibers').

(3) Kripke's second rephrasal strategy is based on the thesis that, when we report an intuition with ' $\text{Possibly, } R_1 \neq R_2$ ', often the true thing we have in mind is strictly and literally reported with ' $\text{Possibly, } D_1 \neq D_2$ ', where ' R_1 ' and ' R_2 ' are names and ' D_1 ' and ' D_2 ' are descriptions. But Kripke, of all people, should not be proposing that, when we make use of a proper-name sentence in ordinary conversation (even if the sentence happens to be of the form ' $\text{Possibly, } R_1 \neq R_2$ ', we have in mind something *descriptive*. After all, the situation is phenomenologically and behaviorally indistinguishable from situations in which we have in mind something *nondescriptive* (as, for example, when Kripke asserts his well-known thesis 'If Hesperus = Phosphorus, then it is not possible that Hesperus \neq Phosphorus'). For Kripke to deny this would be ad hoc and implausible. Hence, the rephrasal strategy itself is implausible. In fact, it can be shown that this rephrasal strategy does not even accomplish the goal of breaking the stalemate between our apparently conflicting pro- and anti-SE intuitions. For one can wield it so as to sustain the original force of the latter and to deflate the original force of the former, thereby rendering our pro-SE intuitions consistent with traditional anti-SE. The following recipe provides one way of doing this. Adopt the traditional description theory of names. Hold that names occurring in reports of anti-SE intuitions are being used strictly and literally and that they express *nonrigid* descriptive content. Hold that names occurring in reports of pro-SE intuitions are not being used strictly and literally and that they are being used to express *rigid* descriptive content.¹³ (For example, this

¹² "Materialism without Reductionism: What Physicalism Does Not Entail," in Block, pp. 67–106, esp. p. 84.

¹³ The following illustrates this three-step recipe in the case of modal intuitions concerning water. (i) Hold that 'water' is synonymous to (something like) 'the

rigidity could be the result of implicitly understood actuality operators.) The rephrasal strategy can thus be used to affirm anti-SE just as effectively as it can be used to affirm pro-SE. Hence, the stalemate is not broken.

In fairness, Kripke tells us that this rephrasal strategy does not always take the form of replacing a statement's rigid designators $\lceil R_1 \rceil$ and $\lceil R_2 \rceil$ with associated definite descriptions $\lceil D_1 \rceil$ and $\lceil D_2 \rceil$. He states that his "general paradigm is to redescribe both the prior evidence and the statement qualitatively and claim that they are only contingently related" (*op. cit.*, p. 143). But the above criticism carries over *mutatis mutandis* to the proposed "qualitative redescrptions" of other statements that report anti-SE intuitions. Specifically, one can sustain the original force of our anti-SE intuitions and, by deft use of actuality operators, always deflate the original force of our pro-SE intuitions, thereby rendering them consistent with traditional anti-SE. So, the second rephrasal strategy, in its full original form, fails to break the stalemate in favor of SE. Of course, Kripke is onto something with his talk of "qualitative redescription of the prior *evidence*." The sound idea is this: it is possible for people in qualitatively the same epistemic situation as ours to be thinking and talking about different things from those we are thinking and talking about. But this is precisely the idea that drives the first rephrasal strategy. "Qualitative redescrptions of the original *statement*" are completely beside the point; indeed, they only create an opening for anti-SE to perpetuate the stalemate.

The conclusion should now be plain. Breaking the stalemate requires dropping the second rephrasal strategy and turning to (some form of) the first. When this is done, the stalemate seems to be broken in favor of SE. According to this strategy, when one wishes to deflate the force of an intuition reported with \lceil Possibly A \rceil , one rephrases that report with \lceil It is possible that a population of speakers in an epistemic situation qualitatively identical to ours would make a true statement by asserting \lceil A \rceil with normal literal

clear thirst-quenching stuff'. (ii) Deem the anti-SE intuition expressed by 'Possibly, water \neq H₂O' to be true, correctly reported, and literally synonymous to the narrow-scope reading of 'Possibly, the clear thirst-quenching stuff \neq H₂O'. (iii) Deem the pro-SE intuition uncritically reported with 'Possibly, there is a twin earth such that . . . the clear thirst-quenching samples on the twin earth are not samples of water and, accordingly, water \neq the clear thirst-quenching stuff' to be true but incorrectly reported; it is correctly reported with (something like) 'There is a possible but nonactual twin-earth world such that . . . the stuff that in the actual world is clear and thirst-quenching \neq the stuff that in the possible but nonactual world is clear and thirst-quenching'. This rephrasal is entirely consistent with traditional anti-SE.

intent'. Because our anti-scientific-essentialists are traditional internalists, they are committed to holding that the meaning of 'A' cannot differ across populations of speakers in qualitatively identical epistemic situations. Accordingly, they must hold that the rephrasal entails the original report. In particular, when the original intuition seems to have a pro-SE force, they are committed to holding that the rephrasal has that force as well.¹⁴ By contrast, scientific essentialists are not traditional internalists, so they are free to hold that the meaning of 'A' can differ across populations of speakers in qualitatively identical epistemic situations. So when the original intuition seems to have an anti-SE force, they are free to hold that that force is deflated upon rephrasal. The stalemate is thus broken in their favor.

The conclusion is that the cogent defense of SE rests on the first rephrasal strategy, not the second. But this conclusion shows that Kripke's defense of the modal argument (and the vast philosophical literature on Kripke's defense) is off the mark.

IV. THE MODAL ARGUMENT REFORMULATED

Consider the cogent SE argument that being water = being H₂O. The argument consists of two steps. First, pro-SE intuitions supporting the identity are elicited: for example, the twin-earth intuition regarding water and XYZ. Second, it is shown that the (first) rephrasal strategy can be used to deflate the force of our anti-SE intuitions but that, when anti-scientific-essentialists attempt to use it to deflate the force of our pro-SE intuitions (i.e., the intuitions elicited in step one), they fail. Because both steps evidently succeed, one may conclude that SE holds for 'water'. To show that it also holds for mental expressions like 'pain', the identity theorist needs to go through both steps.

But the first step fails for expressions like 'pain'.¹⁵ To see this, consider the 'pain' analogue of the original twin-earth argument.

¹⁴ Consider the pro-SE intuition reported with 'Possibly, there is a twin earth such that . . . the clear thirst-quenching samples are not samples of water'. Anti-scientific-essentialists must hold that this intuition is true but incorrectly reported; it is correctly reported with 'It is possible for there to be a population of speakers in an epistemic situation qualitatively identical to ours who would make a true statement by asserting "There is a twin earth such that . . . the clear thirst-quenching samples are not samples of water" with normal literal intent'. But given their internalism, our anti-scientific-essentialists must hold that such a population of speakers would mean what we mean with 'The clear thirst-quenching samples are not samples of water'. This commits our anti-scientific-essentialists to holding that the rephrasal has the same pro-SE force as the original report.

¹⁵ Some people seem to believe that, whenever something is open to any kind of scientific study, SE automatically holds for it. But this just begs the question in the text. For a discussion of why and how the scope of SE must be circumscribed, see my "The Philosophical Limits of Scientific Essentialism," *Philosophical Perspectives*, 1 (1987): 289–365. My *Philosophical Limits of Science* (New York: Oxford, forthcoming) gives a sustained discussion of these matters.

Suppose that on earth all and only creatures in pain have firing C-fibers. Suppose that upon traveling to a twin earth, we find creatures that are macroscopically indistinguishable from the creatures on earth. For example, our human Doppelgänger have behavioral "input/output functions" that are the same as ours. Their "pain" behavior is exactly like ours. They utter "sentences" that, if they were English sentences, would attest to the dreadfulness of pain. Their "dentists" inject something they "call" 'anesthetic'. Their "torturers" are effective in eliciting "confessions." And so on. It turns out, however, that, whereas our pain—and our attendant pain behavior—co-occurs with the firing of C-fibers, the "pain" behavior of the creatures on twin earth co-occurs instead with the firing of C₂-fibers (which, unlike C-fibers, are not composed largely of hydrogen, oxygen, and carbon, but rather of X, Y, Z, and W). Now would we say that these creatures are in pain on the indicated occasions? Well, to be sure, we would not be certain that they are in pain; macroscopic behavioral criteria never entail that a mental predicate applies. Nevertheless, we would have very good evidence that they are in pain. But this is not the point. The point is that it would not be counterintuitive to say that they are in pain. Note the contrast. It would be counterintuitive to say that samples of XYZ on twin earth are samples of water. This intuition is the very foundation of the SE argument concerning water. The analogous intuition concerning pain is simply missing.¹⁶ Accordingly, the foundation of the argument that SE generalizes from physical expressions like 'water' to mental expressions like 'pain' is simply missing. Without this foundation, the thesis that SE does so generalize is nothing but an article of faith.

This conclusion may be used to win a positive result. We have seen that in the case of terms like 'water' there is a pending stalemate between pro- and anti-SE intuitions and breaking the stalemate in favor of SE requires that one apply the rephrasal strategy to reconcile the apparently conflicting intuitions. Now, nearly all philosophers wish to rely evidentially on intuitions (this includes both parties in our debate over the soundness of the modal argument). As a consequence, we are committed to the technique of resolving

¹⁶ Someone might respond that the pain intuition is missing and the water intuition is not because we do not seriously accept the pain/C-fiber correlation whereas we do the water/H₂O correlation. But this cannot be the explanation, for we have water intuitions associated with hypothetical correlations which we actually reject. E.g., we have the intuition that, if our chemists had been deceived and all and only water samples here on earth were really composed of ABC (\neq H₂O) and if there were a twin earth . . . , then XYZ (\neq ABC) samples on the twin earth would not be samples of water.

pending stalemates by means of the rephrasal strategy. Otherwise, we would be forced to admit that a huge number of our intuitions are mistaken and, hence, the evidential weight of intuitions is questionable. But in the absence of a pending stalemate, there is no requirement to subject our intuitions to the rephrasal strategy; absent a pending stalemate, the presumption is that our intuitions are correct as reported. We have just seen, however, that pro-SE intuitions concerning mental expressions like 'pain' are missing. At the same time, we have a wealth of traditional anti-SE intuitions concerning expressions like 'pain'. So the sort of pending stalemate that threatens in the case of intuitions concerning expressions like 'water' is absent in the case of intuitions concerning expressions like 'pain'. Hence, the presumption is that the latter intuitions are correct as reported. But one of these intuitions is the key modal premise in the weak modal argument—that is, the intuition reported with 'It is possible for there to be something that has pain but lacks firing C-fibers'. Therefore, the presumption is that this premise is correct as reported and, hence, that the argument goes through without further ado.

This is the first step in my defense of the weak modal argument. Now the second. Suppose that the above assessment is mistaken and that there is a requirement to subject the argument's key modal intuition to the rephrasal strategy. Identity theorists who would undermine the argument by invoking SE must grant that this modal intuition is true but incorrectly reported and that, when the initial report is subjected to the rephrasal strategy, the resulting rephrasal is true. According to the rephrasal strategy, our true intuition is correctly reported by (something like) 'It is possible that there could be a population of speakers in an epistemic situation qualitatively identical to ours who would make a true statement by asserting "Something has pain but lacks firing C-fibers" with normal literal intent'. If there were a population of speakers in an epistemic situation qualitatively identical to ours, they would use 'something' to mean something; 'has' to mean has; 'lack' to mean lack; 'pain' to mean pain. (If you doubt the last claim, replace 'pain' throughout with the demonstrative 'this', and let it be understood that we are using 'this' for a certain salient phenomenal quality to which each of us is attending during our present experiences of sharp pain. Then our counterparts in a population of speakers whose epistemic situation is qualitatively identical would have to be using 'this' for an identical quality.¹⁷) But people in a qualitatively identical epistemic

¹⁷ Formulated this way, my argument counts against David Lewis's "An Argument for the Identity Theory," this JOURNAL, LXIII, 1 (January 6, 1966): 17-25.

situation could use 'firing C-fibers' to refer to something other than firing C-fibers. (This is the echo of Boyd's criticism of Kripke.) For example, they could inhabit a possible world in which the following hold: except for the envisaged speakers, all apparently sentient beings are nonconscious automata; the speakers' term 'firing C-fibers' refers paradigmatically to a silicon-based process that their scientists commonly observe in the "brains" of these automata; this process never occurs in any being that is conscious (i.e., it never occurs in any of the speakers). In view of such possibilities, the rephrasal of the key modal intuition undermines the original force of that intuition.

This negative outcome is only temporary, however. The argument can be reformulated with an alternate modal premise whose force survives rephrasal. The idea is to find a substantive necessary condition for firing C-fibers which is specified in terms whose basic semantic properties would be the same for any population of speakers whose epistemic situation is qualitatively identical to ours. We already know that there are some terms like this. For example, any population of speakers whose epistemic situation is qualitatively identical to ours would use 'has' to mean has and 'pain' to mean pain. We are looking for a substantive necessary condition for having firing C-fibers that can be described in terms with this kind of semantic stability.

Surely, there is a necessary condition that fills the bill. Our larger goal is to show that SE does not generalize from items like water to items like pain. Suppose for reductio that it does. Then certainly it would generalize from items like water to items like C-fibers. Just as water has an essence discoverable by chemists, C-fibers would have one discoverable by cell biologists. Now, C-fibers are phylogenetically comparatively old unmyelinated neurons with a distinctive morphology: an item would not *be* a C-fiber unless it had certain specific kinds of nonconscious functionally related parts. How many such parts must C-fibers have? The count may be performed on a very small scale—perhaps even at a level of macromolecules. So the count can be very large, far exceeding anything the ancients would have dreamt. Suppose it is at least 74,985,263. Then it is necessary that an item is a C-fiber only if it has 74,985,263 or more nonconscious functionally related parts. In turn, it is necessary that a being *x* has firing C-fibers only if *x* has some parts that have 74,985,263 or more functionally related nonconscious parts. This necessary condition has the desired semantic stability: any population of speakers in an epistemic situation qualitatively identical to ours would use 'has some parts that have 74,985,263 or more functionally related nonconscious parts' the way we do. By '74,985,263',

they would mean 74,985,263; by 'function', function; by 'relation', relation; by 'non', non; by 'conscious', conscious; and so on.¹⁸

Using these ideas, we may reformulate the weak modal argument thus:

- (1) It is possible that there could be a being who is in pain but lacks parts that have 74,985,263 or more functionally related non-conscious parts.
- (2) If the property of being in pain = the property of having firing C-fibers, then necessarily, for all x , if x is in pain, x has firing C-fibers.
- (3) Necessarily, for all x , if x has firing C-fibers, x has some parts that have 74,985,263 or more functionally related nonconscious parts.

Therefore, the property of being in pain \neq the property of having firing C-fibers.

The argument is valid. (2) is undeniable. Because (3) is supplied by SE, we are entitled to accept it. After all, the threat to the original modal argument came from SE. So the issue comes down to (1), which seems intuitive. When the rephrasal strategy is applied to (1), we get 'It is possible for there to be a population of speakers in an epistemic situation qualitatively identical to ours who would make a true statement by asserting "There is a being that is in pain but lacks parts that have 74,985,263 or more functionally related non-conscious parts"'. But such populations of speakers would mean by this sentence what we mean, as already indicated. Hence, the rephrasal entails the original report. So, given that scientific essentialists are committed to the truth of the rephrasal, they are committed to the truth of the original report. Thus, they are forced to accept (1) and, in turn, this new modal argument.

Let us summarize. First, the intuitions needed to extend SE from terms like 'water' to terms like 'pain' are simply missing. Therefore, there is no apparent conflict between pro- and anti-SE intuitions regarding such terms. In view of this, there is no evident requirement to subject the original argument's key modal intuition to the rephrasal strategy. Absent an apparent conflict among our intuitions, the presumption is that the original argument is sound as it stands. Second, even if this assessment is mistaken, the argument

¹⁸ Even if these terms meant something different, their meanings would be very close to the English (this is all that is needed to construct a possibility that is inconsistent with the thesis that being in pain = having firing C-fibers). The indicated terms are thus fundamentally different from terms like 'water', 'tree', 'vat', 'brain'; considerations like those raised by Putnam ("Brains in a Vat," *Reason, Truth and History* (New York: Cambridge, 1981), pp. 1-21) therefore do not apply.

can be reformulated so that (I)—its key modal premise—retains its original force even when subjected to the rephrasal strategy. This revised argument withstands the SE critique “twice-over.”

We thus have a successful modal argument against the thesis that being in pain = having firing C-fibers. Presumably, for any finitely stateable first-order physical state S_i , the argument could be adapted to refute the thesis that being in pain = being in S_i .

Despite this result, we have not yet ruled out the thesis that being in pain has an infinitary (or indefinitely long) definition-by-cases. So the following sort of property identity is still in the running: being in pain = being in S_1 or S_2 or S_3 Advocates of the multiple-realizability argument would not be troubled by this prospect, for they are prepared to invoke auxiliary assumptions to exclude this kind of property identity (see section one). But such auxiliary assumptions are quite problematic.

To refute this sort of property identity, one must turn from the above weak modal argument to a strong modal argument. Notice that the property of having a multiplicity of functionally related nonmental parts is a necessary condition of every physical condition — S_1, S_2, S_3, \dots —of the sort identity theorists would entertain. Suppose the first premise in our weak modal argument is replaced with the following: it is possible for there to be a being who feels pain but does not have a multiplicity of functionally related nonmental parts. The resulting argument is valid, so its correctness comes down to the truth of this premise. Many people—if they set aside philosophical biases—find this new modal premise intuitive. Moreover, the intuition supporting it survives both stages of aggressive SE attack. First, there is no contrary twin-earth intuition to compete with it. So the presumption is that it is immune to SE worries. At the second stage, the intuition also fares well. For, when our report is subjected to the rephrasal strategy, it retains its original force. But it must be admitted that many of us fail to find this strong modal premise flat-out obvious; if we accept it, we do so with hesitation. For this reason, we are not entitled to conclude with any conviction that the identity thesis in its full generality is false.

The problem, to repeat, is that the new premise is controversial. Is there any prospect of settling whether that premise is false or true? Yes. After all, in the case of the weak modal argument, we were able to establish that it is possible for there to be a being that feels pain but lacks parts that have 74,985,263 or more functionally related nonconscious parts. Virtually everyone (even the identity theorist) has the intuition that this is possible. This intuition is as vivid as those invoked in other successful philosophical arguments. (For ex-

ample, the intuitions invoked to show that it is possible to have justified true beliefs that are not knowledge or, indeed, those invoked to establish SE—twin earth, cat-looking robots, gold-appearing compounds, nonidentical qualitative duplicates.) Moreover, this intuition retains its original force upon SE rephrasal, so it is the kind of intuition that must be accepted according to SE procedure. It is thus possible for there to be beings (i.e., you and I) who are able to answer a priori the question of whether it is possible for there to be a being that feels pain but lacks parts that have 74,985,263 or more functionally related nonconscious parts. Let us grant that, by contrast, you and I are not able to answer a priori the question at issue in the strong modal argument (i.e., the question of whether it is possible for there to be a being that feels pain but lacks a multiplicity of functionally related nonmental parts). Let us also grant that the answer to this question might well differ from the answer to the first question. Nevertheless, the two questions are otherwise quite alike: they each concern the possibility of pain in the absence of certain nonmental properties. Moreover, those properties—and the questions themselves—are expressed in the very same sort of semantically stable vocabulary, which retains its original force upon SE rephrasal. Given that it is possible for there to be beings (you and I) who are able to answer the first question a priori, then unless a specific barrier is identified, uniformity supports the presumption that it should also be possible for there to be beings (though perhaps wildly more intelligent than you and I) who are able to answer the second question a priori.

V. THE CERTAINTY ARGUMENT REFORMULATED

The original certainty argument was invalid because it was formally equivalent to a scientific-definition analogue of the paradox of analysis. The above discussion of the modal argument suggests that this difficulty can be avoided by “modalizing” the certainty argument:

- (1) It is possible for there to be a being who knows a priori whether or not it is possible for something to be *F* and not *G*.
- (2) It is possible for there to be a being who knows that he is *F* solely by virtue of the fact that he currently is *F*.
- (3) If it is possible for there to be a being who knows a priori that it is not possible for something to be *F* and not *G*, and it is possible for there to be a being who knows that he is *F* solely by virtue of the fact that he currently is *F*, then it is possible for there to be a being who knows that he is *G* solely on the basis of reason and the fact that he is *F*.
- (4) It is not possible for there to be a being who knows that he is *G* solely on the basis of reason and the fact that he is *F*.

Therefore, it is possible for something to be *F* and not *G*.

This argument is valid. To see why, suppose that it is not possible for something to be *F* and not *G*. Then, (1) would imply that it is possible for there to be a being who knows a priori that this is indeed not possible. From this, (2), (3), and *modus ponens*, it follows that it is possible for there to be a being who knows that he is *G* solely on the basis of reason and the fact that he currently is *F*. But this contradicts (4). So the supposition is false, and the desired conclusion follows: it is possible for something to be *F* and not *G*.

If '⌈*F*⌋' is 'is in pain' and '⌈*G*⌋' is 'has a multiplicity of functionally related nonmental parts', this conclusion is just the first premise of the strong modal argument, namely, the premise that it is possible for there to be a being that is in pain but lacks a multiplicity of functionally related nonmental parts. Since the remainder of that argument is unproblematic (even to identity theorists), its conclusion would follow: the identity thesis in its full generality would be false. Thus, the reformulated certainty argument may be thought of as a way of establishing the first premise of the strong modal argument without having to rely on a controversial modal intuition.

Since this reformulated certainty argument is valid, its correctness turns on the truth of the premises. But each one has good reasons supporting it. With '⌈*F*⌋' and '⌈*G*⌋' as indicated,¹⁹ (1) is: it is possible for there to be a being who knows a priori whether or not it is possible for a being to be in pain and not have a multiplicity of functionally related nonmental parts. This premise, however, is the point defended at the close of the previous section when we compared the weak and strong modal arguments. The weak modal argument is concerned with the question of whether it is possible for a being to be in pain and not have parts that have 74,985,263 or more functionally related nonconscious parts. The strong modal argument is concerned with the question of whether it is possible for a being to be in pain and not have a multiplicity of functionally related nonmental parts. Given that it is possible for there to be a being (for example, you or I) who is able to know the answer to the first question a priori, what could block the possibility of a being (perhaps wildly more intelligent) who is able to know the answer to the second question a priori? At present we have no reason to think that there

¹⁹ If '⌈*G*⌋' were 'has firing C-fibers' or 'is in *S*₁' . . . , the resulting premise (1) would run afoul of an echo of Boyd's point. This problem is circumvented with the semantically stable predicate 'has a multiplicity of functionally related nonmental parts' which provides just a necessary condition for these physical predicates.

The analogous reformulation of the Nagel-Jackson argument would instead need semantically stable predicates that provide sufficient conditions for 'has firing C-fibers', But there are none; see the "Afterword."

exists a barrier to the possibility of this a priori knowledge.²⁰ So, unless we establish that there exists a barrier, we must either accept the possibility of this a priori knowledge, or flaunt uniformity considerations and accept an otherwise avoidable mystery.

This leaves us with (2), (3), and (4). (2) is intuitively obvious. It says that one can know that one is in pain solely by virtue of the fact that one is currently in pain.²¹ (3) is also intuitively obvious. It says that, if it is possible for someone to have a certain kind of a priori knowledge and someone to have a certain kind of introspective knowledge, then it is possible for someone to have both kinds of knowledge and to use them to derive an immediate logical consequence of them. (4) is similarly unproblematic when understood as intended. When I say that a person uses only reason and the fact that he is currently in pain, I mean to rule out that the person has had any prior exposure to physical theories, their instruments, or their terminology.²² I also mean to rule out that the person is making any evidential use of the following: sensory facts besides the fact that he is in pain, memories of sensory facts, previously learned

²⁰ Epistemology has uncovered only two relevant potential barriers to a priori knowledge of whether a proposition is possible. The first is SE. But we have shown that the terms used to state the proposition at issue in premise (1) are immune to both stages of the SE argument, and so we have no reason to think that SE blocks the possibility of a priori knowledge of whether that proposition is possible. The second kind of potential barrier results from limitations on intelligence: if a being's intelligence is too low, the being will be unable to obtain a priori knowledge of various possibilities. But we have no reason to think that there are necessary limitations on intelligence which block the mere possibility of a being with the a priori knowledge at issue in premise (1). On the one hand, the sort of question at issue (i.e., whether pain without a multiplicity of functionally related nonmental parts is possible) does not seem to be the sort of question which would require infinitary intelligence (e.g., for doing infinitary proofs, infinitary computations, etc.). Some finite level (though perhaps wildly beyond ours) ought to suffice. But, intuitively, for any finite level of intelligence, it is possible for some being to be that intelligent. (This intuition is expressed in semantically stable terms, so SE is committed to accepting it.) On the other hand, even if infinitary intelligence were required (e.g., for checking proofs with infinitely many premises, etc.), that should not matter, for we have no reason to doubt the possibility of infinitary intelligence. There are no intuitions that go against this possibility; if anything, intuitions support it. And there are no good arguments against it. (It is far weaker than the possibility of an omniscient, omnipotent, benevolent God; so familiar arguments against the existence of God—e.g., the argument from evil—have no bearing.) SE certainly provides no argument against this possibility.

²¹ I use 'knows' not 'is certain', so there is no commitment to Cartesian infallibility. Instead of 'knows', I could use the still weaker phrase 'judges truly and is able to justify'.

²² Churchland, *op. cit.*, claims that people with prior exposure to physiological theory and its terminology could have introspective knowledge of their own brain states. I find this implausible. In any case his claim has no bearing on (4), for as it is intended (4) rules out such prior exposure.

empirical theories, testimony of others, and so forth. Understood this way, (4) is intuitively obvious: if the evidential resources are restricted in all these ways, plainly the person could not on that slender basis know that he possesses a multiplicity of functionally related nonmental parts. When these points are fully in view, even people in the grips of the identity theory must accept (4). For its denial is tantamount to holding that, from the sole fact that one is in pain, it is possible to give an otherwise a priori argument for the existence of the external world and one's bodily station in it. This flies in the face of the main lessons of modern epistemology.²³

The argument therefore seems to go through. We used the weak modal argument to give an a priori defense of the first premise of our certainty argument—namely, it is possible for there to be a being (perhaps wildly more intelligent than ourselves) who is able to know a priori whether pain without a multiplicity of functionally related nonmental parts is possible. This premise, together with the other three premises, yields the conclusion that pain without a multiplicity of functionally related nonmental parts is indeed possible. Because the argument is wholly a priori, it turns out that we ourselves are beings who can know a priori that pain without a multiplicity of functionally related nonmental parts is possible.

As indicated, this bit of knowledge is the first premise of the (otherwise uncontroversial) strong modal argument. By thus weaving together the weak modal argument, the certainty argument, and the strong modal argument, we arrive at a refutation of the identity thesis in its full generality.

VI. AFTERWORD

This argument is consistent with certain versions of functionalism and also with the thesis that there are first-order physical properties that are logically sufficient for mental properties. The Nagel-Jackson knowledge argument is bolder in that, if correct, it would refute

²³ Using only your rational faculties and the empirical fact that you are currently experiencing pain, would you be able to establish that you have a multiplicity of functionally related nonmental parts? No, for it is a matter of real controversy whether or not there could be a being who experiences pain but has no body. If your empirical evidence were restricted to your current pain, you would be unable to rule out the proposition that you had no body. Suppose that you were to become more intelligent. Might there be some formal logical consequence that you had missed which would now come into view so that you could deduce that you have a body? That is implausible. With your greater intelligence, might you now be able to make an intuitive inference that you have a body? That too is implausible. Perhaps with your greater intelligence you could discover a new theory or argument establishing that you have a body. Although we are not in a position to rule out this prospect with certainty, the distinctive pattern of failure in modern epistemology makes this prospect very dim. Moreover, it is implausible that the situation would be reversed by further increases in intelligence.

both functionalism and the sufficiency thesis. Does the above critique of aggressive SE help to salvage the knowledge argument? No. Functionalists may rebut the argument just as before (see note 9). Moreover, even if functionalism were known to be false—say, because certain qualia are totally indistinguishable functionally—knowledge arguments still could not refute the sufficiency thesis. To see why, suppose Mary has experienced one of these qualia *P* and has a semantically stable expression for it. Suppose there is a creature who also experiences *P* but who differs significantly from Mary in appearance, behavior, and physiology. Then, presumably, from the physical facts Mary could not infer a priori that the creature experiences *P*. This would not imply, however, that the creature's physical properties are not logically sufficient for experiencing *P*. For there is an alternate explanation: the a priori inference would be possible only if both the mental property *P* and the relevant physical properties were expressible with semantically stable expressions; but, echoing Boyd, SE shows that such expressions are not possible for the physical properties. (For example, consider a possible world whose matter is structurally and functionally like the matter in the actual world but which differs from the latter in scale. Every semantically stable predicate that would apply to "fibers" composed of the one type of matter would apply to "fibers" composed of the other. The problem is that, if functionalism is false, "fibers" composed of the new kind of matter might fail to be sufficient for *P*.) The fact that SE applies to the creature's physical properties could thus explain why Mary cannot know a priori the (supposed) fact that the physical properties are logically sufficient for *P*. So, *pace* Nagel and Jackson, the existence of unanswerable "open questions" about the "subjective" (even in the presence of all the relevant "objective" knowledge) would not imply that physical properties are not logically sufficient for mental properties.

GEORGE BEALER

University of Colorado