23. QUALIA, KRIPKEAN ARGUMENTS, AND SUBJECTIVITY

EMMETT L. HOLMAN
George Mason University

ABSTRACT. The subjectivity of consciousness is widely regarded as a major stumbling block for materialist theories of mind. In this paper I show how Kripkean arguments against identity theories (Kripke, 1972), and in particular a Kripkean argument against qualia—material property identity developed by Frank Jackson (1980) are a way of highlighting this problem. (And such arguments are not the quasi-historical curiosities they are sometimes pictured as being, because problems confronting functionalism have led to a modest revival of identity theory.) As such, Kripkean arguments are akin to recent discussions of subjectivity by Thomas Nagel (1965, 1974, 1979) and Frank Jackson (1982). I then consider some recent attempts to refute Kripkean arguments or otherwise show that subjectivity is not an insurmountable problem for identity theory. The most promising attempt is one that I myself develop, based on some ideas by Keith Gunderson (1970). But I contend that even it, let alone any of the others, is not without problems. Thus, tentatively, Kripkean arguments against property identity succeed.

The subjectivity of consciousness is widely regarded as a major stumbling block for materialist theories of mind. Exactly what such subjectivity is is often not made clear, but two related attempts to highlight its importance, or so I will argue, are the appeal to so-called "qualia" and the kind of "modal" argument against identity theories originally developed by Saul Kripke (1972). In current debates over the mind-body problem these two are related as follows: The appeal to qualia is used to show that theories which construe the mind entirely as a functional or causal system necessarily leave something out. In view of the difficulties thereby highlighted a few writers have recently proposed that at least a modified form of identity theory be adopted for qualia. So, Patricia and Paul Churchland (1981) have suggested that qualia can be identified with neurophysiological properties so long as we keep in mind that these properties are not the means by which we individuate sensation-types. (Thus is species chauvinism avoided.) This allows a kind of materialistic functionalism in which qualia are the natural kinds which are instrumental in fulfilling various functional roles—keeping in mind that, for different creatures, the same functional role can be fulfilled by different qualia. And Terence Horgan (1984) has proposed a "two part" theory of mind, in which qualia are psychological
state-types which are identical with neurophysiological state-types, and non-qualitative psychological state-types are construed functionally. (And Horgan is apparently willing to live with the limited species chauvinism this entails.) Thus are elements of identity theory introduced to save the day for materialism. But it is precisely such identity theories that Kripkean arguments claim to undermine.

Frank Jackson (1980) has recently developed a form of Kripkean argument, dealing with property identity, which is not vulnerable to the criticisms first leveled against Kripke’s original argument. (See Feldman, 1974, 1980; Lycan, 1974.) After some preliminary remarks in Section I of this paper, I will, in Section II, present Jackson’s argument and then broaden it to show that its real point is, indeed, the apparent inability of materialist identity theories to account for the subjectivity of consciousness. (In fact, I will even suggest that my broadened argument can thereby be used against certain forms of functionalism.) In Sections III-V I consider possible objections to this, including recent objections to Kripkean arguments as well as to related treatments of the problem of subjectivity by Thomas Nagel (closing remarks in 1965; also 1974, 1979) and Frank Jackson (1982). The objection in Section V is, so far as I know, new, being based on an article by Keith Gunderson (1970) which, while not dealing with Kripkean arguments, suggests an intriguing way of doing so. Despite showing the most promise, I don’t think this approach, let alone any of the others I examine, unproblematically refute my or Jackson’s arguments, or otherwise show that subjectivity is not a stumbling block for a materialist identity theory.

In the course of my discussion I will have occasion to speculate on the possible identity of qualia with certain kinds of non-material, as well as material, properties. But it should be remembered that, on the whole, a point of departure of this paper is the apparent failure of purely causal/functional accounts of qualia. So, except when I indicate otherwise, I will not be concerned with properties definable by causal or functional role. (I will relax this restriction only when I address the applicability of Kripkean arguments to functionalism.) Perhaps a good way of delineating the properties I am (mainly) concerned with is as follows: They are all those properties which constitute a natural kind in some logically possible world.

Finally, the kind of materialist identity theory against which I claim Kripkean arguments to be successful is one which understands a material property in terms of Meehl’s and Sellars’ (1956) “physicals”. I.e., a property is material iff it is subsumable under those laws which account for non-living reality. (And since vitalism is not at issue here, we can assume that the laws in question apply to non-minded reality.) There are, of course, weaker senses of “material”. E.g., Meehl’s and Sellars’ “physicals” applies to something iff it can be situated in the spatio-temporal-nomological framework. But most self-styled identity theorists would not be content with a quale’s being material in only this sense.

I

Kripkean arguments address statements of the form

\[ S_i = M_j \]
to see if any of them might be true. Since our concern is only with property identity, we can take the left side designator to be any member of the set, \( \{S_i\} \), of rigid "sentient property" or "qualia" designators, and the right side designator to be any member of the set, \( \{M_j\} \), of rigid "materialistic" property designators. (Here a rigid designator is defined in the usual way as a designator that designates the same entity in all possible worlds in which that entity exists.) For a more specific account of what qualifies a designator for membership in \( \{M_j\} \), something more or less along the following lines should do for our purposes (where the "more or less" applies to the second condition below). First, the designator has a material property as referent. And second, the designator is introduced with the intent of designating, or is known by its users to designate, a material property. Parallel remarks apply for membership in \( \{S_i\} \).

Kripkean arguments are designed to show that no statement of form (I) can be true, but before we consider this some preliminaries need to be dealt with. First, in order for any rigid designator to do its work, it must have its reference fixed in such a way as to indicate to language users appropriate conditions for its use. This is sometimes done by appealing to an actually instantiated manifest property or characteristic which signals the purported presence or activity of the property to be designated. I will call such a characteristic the "reference fixing characteristic" (rfc) of the rigid designator whose reference it thereby fixes. It will be safe to assume that all rigid designators relevant to the topic of this paper have their references fixed by an rfc.

Now if the "purported presence or activity of the property to be designated" is to be signaled to the language user, it must make a difference in his experience. So at the most basic level the rfc of a designator is always a feature of experience. Sometimes it may not be necessary to address matters at such a basic level, but in this paper it will. A consequence of this that I will have occasion to use further on is the following:

\[(A)\] Any member of \( \{S_i\} \) or \( \{M_j\} \) has an rfc which is itself a referent of a member of \( \{S_i\} \).

Sometimes the rfc of a designator may be only contingently associated with the referent. When this is so, we can look upon the reference fixing procedure as follows: There is a "reference fixing description" (rfd) which non-rigidly designates the referent via its connection to the rfc. The connection indicated may be of a very rough and ready sort—say, that the referent causes the rfc in some unspecified way—or it may employ the resources of a carefully worked out theory. The reference fixing then occurs by indexing the rfd to the actual world.

But we need not confine the idea of an rfd just to cases of contingently associated rfc's. It will be convenient to look upon all reference fixing as involving an rfd. When the rfc and referent are identical, the rfd can be taken as "\((\forall x) (x = R)\)" where "R" rigidly designates the rfc. And if there are other ways in which rfc and referent can be necessarily connected (whether or not there are won't matter for the purposes of this paper) the corresponding rfd will reflect that connection.
I have said that rfc's are "manifest". Let me offer a preliminary explanation of what I mean. (More will be said later.) Basically, I intend for this term to capture what philosophers have traditionally meant by a property's being "observable" or "directly observable" or "non-inferentially observable"—and I mean for this to include introspectability. Now I realize there are all sorts of problems here, but I think I can justifiably beg off dealing with them in this paper. My point of departure, recall, is the apparent failure of functionalism to deal with qualia, and the standard arguments which lead to this point (the inverted spectrum argument, e.g.) certainly make implicit use of a kind of introspective manifestness. Furthermore, my understanding of being manifest is compatible with a wide range of views on theory of perception, the theory-ladenness of observation, etc. I can even, up to a point, accommodate the view that radical conceptual or linguistic changes can change the set of properties which are manifest to someone. But there is a point beyond which I cannot go. Since I wish to examine the prospects that materialism can accommodate qualia without explaining them away, I must reject the view that the appropriate conceptual or linguistic changes can lead us to truly conclude that there are no actually instantiated qualia or rfc's for qualia designators.

II

The question, now, is whether or not statements of form (I) are true. To make things slightly less abstract, suppose we let the sentient property designator be "having a sensation of P" (hereafter "having S(P)"). (We could just as well have chosen "being in S(P)", or "being a S(P)"). The materialist designator will be "M". Our question, then, is addressed to

\[(I) \text{ having } S(P) = M\]

Adapting Jackson's Kripkean argument to this case, it would go as follows: Since both designators in (I) are rigid, it is true only if it's necessarily true. But it doesn't seem to be necessarily true. For any pair of designators of the kind in question, it seems one can imagine possible worlds in which they have different extensions. Hence, either (I) is not true, or there is another explanation of its seeming non-necessity. The only other explanation available, according to Jackson, (I will later consider other alternatives he doesn't acknowledge) is that the sensation designator's rfc is contingently related to the common referent of the two designators. Being so contingently related to the referent, it cannot be identical with it, and since the referent can be any material property, the rfc is non-material. So one way or another, we must admit that there is an actually instantiated non-material property. As far as the issue of materialism is concerned, we might as well say that it is (the property) having $S(P)$. If this argument works, it undermines the materialist pretensions of both Horgan's and the Churchlands' accounts of qualia which were cited in my introductory remarks. Both parties accept, either explicitly or implicitly, the necessary assumption—namely, that qualia designators are rigid designators of actually occurring states (in Horgan's case) or actually instantiated properties (in the Churchlands'). Such, I submit, is the gist of the admission that qualia cannot be explained away or construed causally/functionally. We can then go on and accept Horgan's
identification of qualia with neurophysiological states, but still point out that Jackson's argument commits him to property dualism. And we can also point out that, as far as Jackson's argument is concerned, the Churchlands gain nothing by distinguishing between the havings of sensations and qualia (And henceforth I will not do so.)

I now want to take a closer look at Jackson's argument to draw out its full import. This will require further elaboration on the idea of a property's being manifest. For now I will take as primitive the concept: x is manifested to (roughly, is observed by) a cognitive being, or cognizer. (But later I will have something to say about this concept.) The variable will be allowed to range over both properties and individuals. For now our interest is in properties. But certainly, whenever a property is manifested to a cognizer, it is so manifested qua being instantiated by some individual, and in (D4) below I will have to put it in this way. (An intellectual intuition, Platonic style, of the property per se is of course not the point here.) Usually, however, a simpler locution will suffice. But I wish to leave it open that a property, qua being instantiated by an individual, i, can be manifested to a cognizer, c, without i's itself being manifested to c. It may be questioned whether or not this is intelligible. But allowing for this is not essential to Kripkean arguments; rather it is a concession to the materialist opponent, as we will see in Section V. Until then, we can put this rather odd idea on the back burner.

So in dealing with the manifestation of properties, I will usually talk simply in terms of a property's being manifested to a cognizer. I begin below by defining the term "manifest" relative to types of cognizers and logically possible worlds. A "cognizer" is anyone (or anything) that is capable of acquiring information or would-be information about himself and his environment. A type of cognizer is defined in terms of those features of a cognizer that are relevant to acquiring such information. For human-type cognizers this would include physiological features. But there are logically possible cognizers--such as disembodied minds--for which this is not so. Furthermore, in accordance with my earlier remarks, I have no objection if the reader wishes to include a cognizer's theoretical presuppositions among the determining factors of his type. This would mean, of course, that an individual cognizer could change type as scientific revolutions come and go.

I now offer the following definitions.

(D1) Property P is manifest to K-type cognizers in logically possible world $W = \sigma$. There are K-type cognizers in W, and it is compatible with the laws of nature of W that P be manifested to K-type cognizers in W.

(D2) Property P is manifest to cognizer c in possible world $W = \sigma$. c exists in W, and is a type of cognizer to which P is manifest in W.

(D3) Property P is manifest in W There is a possible world, W, and a cognizer type, K, such that P is manifest to K-type cognizers in W.

(D4) Property P is privately manifest in W There is a possible world, W, and a cognizer type, K, such that P is manifest to K-type cognizers in W.
EMMETT L. HOLMAN

sible world, \( W \), if \( x \) instantiates \( P \) in \( W \), and \( P \), qua being instantiated by \( x \), is manifest to \( c \) in \( W \), then \( x = c \).

(D5) Property \( P \) is subjective \( \equiv \) Either (a) \( P \) is privately manifest, or (b) for any \( x \), and any possible world, \( W \), \( x \) instantiates \( P \) in \( W \) only if \( x \) instantiates some privately manifest property in \( W \).

(D6) Property \( P \) is objective \( \equiv \) \( P \) is not subjective.

Let me make a few comments on these definitions. First, I will assume that anything (D5) adds to (D4) is relevant only to unmanifest properties, if there are any. So I am assuming that a subjective property is either privately manifest or not manifest at all. If the reader thinks that this is not already implied in (D5), she can add it. Second, it follows from these definitions that all material properties are objective. But the converse does not hold.

One more definition is necessary.

(D7) "0" is used in world \( W \) as an objectivist rigid designator \( \equiv \) (a) "0" is used in \( W \) to rigidly designate an objective property, and (b) "0" is introduced in \( W \) with the intent of designating, or is known by its users in \( W \) to designate, an objective property.

This, of course, is a generalization of the definition of a materialist designator given in Section I. (And we can hereafter assume that (A) in Section I has been similarly generalized.)

Let us now consider a designator, "0", which is used as an objectivist designator in some possible world, \( W \), and a qualia designator, "having \( S(P) \)" with rfc \( r_s \). I then submit that

(I2) having \( S(P) \) = 0

would not seem to be a necessary truth to any cognizer in \( W \). So by Jackson's line of reasoning, either it is not true, or \( r_s \) is not identical with the referent of "0". (Again, other alternatives will be considered later). Let's consider the latter case. Since the referent can be any objective property, \( r_s \) is subjective. And since \( r_s \) is an rfc, it is manifest. Furthermore, since a subjective property is manifest only if it is privately manifest, \( r_s \) is privately manifest. Suppose, on the other hand, that (I2) is false. Then having \( S(P) \) is subjective, which means either that it's privately manifest, or is logically linked to some instantiated property that is. In any case, then, there is not only an instantiated property that is subjective, but one that is privately manifest. And since this instantiation occurs in any world in which having \( S(P) \) is instantiated, it occurs in the actual world. So as far as an "objectivist" theory of mind is concerned, we might as well say that having \( S(P) \) is the privately manifest, hence subjective, property.

What this argument demonstrates is the importance of the objective-subjective dichotomy, and the fact that qualia fall into the latter category and are, moreover, privately manifest. I maintain that this is also the real point of Kripkean arguments generally, rather than materi-
alism vs. non-materialism per se. If so, then it is also clear that what is at stake as far as materialism is concerned is indeed physical rather than physical, since I see nothing to preclude a subjective property from being the latter.

If I am right about all this, it raises an interesting possibility. I see no obvious reason why the subjective-objective distinction can't be construed so as to apply to causal and functional properties in addition to natural kind properties. If so, then the causal/functional properties would certainly fall into the objective category, and we could use my expanded Kripkean argument to defeat some forms of functionalism. It is no barrier to such a move that many functionalists regard folk psychological designators to be non-rigid. As Jackson (1980) points out, a non-rigid designator can always be converted to a rigid designator by prefixing it with "being", "being in" or "having". We could then run the argument for those converted designators. What does provide a barrier to a sweeping refutation of functionalism in this way is that some functionalists present their theory as an analysis of psychological designators. (See Lewis, 1966, 1972 and Armstrong, 1968. Also Smart, 1959, can be looked upon as an early version of analytical functionalism.) Such functionalists would maintain, contra Kripkean intuitions, that, if we pay close enough attention to the meanings of the terms used in the appropriate identity statements, those statements do seem to be necessary truths. But there is also "non-analytical" functionalism. (Fodor, 1968; Harman, 1973; Putnam, 1971.) Against this my expanded Kripkean argument does, I think, work.

I can think of several objections that may be raised against the argument of the foregoing paragraph, but that's a tangent I don't want to pursue. Nothing else that transpires in this paper depends on the argument, so if the reader wishes to, she can forget it. What I definitely do want to carry away from my expanded Kripkean argument is the focus it places on the issue of subjectivity. Now the idea that subjectivity is the real issue underlying the mind-body problem is a major theme of much of Thomas Nagel's writings, and also of another recent article by Frank Jackson (1982). Nagel can perhaps be variously interpreted, but as I see it he addresses two different kinds of subjectivity in his writings, and one of them certainly seems to correspond to mine. If one can't know what it's like to be a bat simply in virtue of having a scientific account of bat perception, then it appears that experience is subjective in the way I have defined it. And Jackson is certainly making the same point in his 1982 article, his quibbles with Nagel to the contrary notwithstanding. (See n. 9 and 10.) Jackson imagines an individual, Mary, who from birth is confined to an environment in which she visually experiences only shades of gray. Despite these constraints, she grows up to become an outstanding neurophysiologist, knowing everything there is to know about the neurophysiology of visual experience. But, Jackson insists, having such neurophysiological knowledge would not by itself enable her to know what it's like to have visual experiences of colors other than shades of gray.

If the mind-body problem really does reduce to the problem of subjectivity, then it would seem that nothing is to be gained by adopting substance dualism. This, in fact, is the conclusion Nagel draws. If there is a mystery about how subjective properties can occur in the objective world, then it is made no less of a mystery by having these properties inhere in separate, non-material substances. This point is...
somewhat kindred to the main point which emerges from my broadened Kripkean argument—qualia are not only not material, they are not objective at all. (It's not quite the same point, however, since I suppose someone could postulate the existence of a substance which has only subjective properties. But this still wouldn't solve the "mystery" of subjectivity, if it is regarded as a mystery in the first place.)

III

The foregoing observations can be used to respond to some things said by the Churchlands which may seem to undermine Kripkean arguments.

One criticism appears in two recent articles by Paul Churchland (1985a, 1985b). It is directed primarily at Jackson's 1982 argument, but in one of the articles (1985b) Churchland suggests that it can be used against Kripke as well. The criticism takes the form of a reductio, designed to show that Jackson's argument proves too much. If Mary the sensorily deprived neurophysiologist could not know what it's like to have certain experiences, neither could a similarly deprived "ectoplasmologist"—i.e., "scientific substance dualist". But all this shows is that nothing is to be gained by adopting substance dualism, and I have already conceded that. But that doesn't refute a subjective-objective property dualism.

At one place in Churchland & Churchland (1981) the authors address the objection that qualia don't seem to be material by saying that awareness of qualia is opaque. As a way of sustaining an identity theory, this reply is defeated by Jackson's Kripkean argument, let alone mine. But it does suggest another prima facie promising form of materialism.

I will assume that the Churchlands mean that introspection is insufficient to tell us either that qualia are material or that they aren't. I will then turn this into an ontological point by proposing that qualia are determinable properties that have both material and non-material determinates. The materialist is then free to say that all actually instantiated determinates are material. At first glance the idea of such a determinable property may seem rather hard to swallow. But it should be no more so than that of a kind of awareness that cannot distinguish between a property's being material or non-material, and so far as I know nobody objects to that.

This proposal has a lot to be said for it. It acknowledges that there are qualia which are not merely causal or functional properties. It avoids the "species chauvinism" objection, since there may be many different actually instantiated determinates of each of the determinable in question. And it appears to avoid Kripkean arguments, since, having non-material determinates, the properties in question are not material.

Now if Kripkean arguments were just arguments against materialism, I would regard this proposal as worth pursuing. But, alas, as we have seen that, properly understood, Kripkean arguments are arguments against any theory which maintains qualia to be objective; and the theory just proposed is such a theory. So it doesn't work, and therefore neither does the Churchlands' "opaqueness" ploy.
Perhaps this ploy was intended as a set-up for a move Paul Churchland makes in 1985a. Applying his favorite theme that theory can transform experience, (see Churchland, 1979, chapter 2) he suggests that sufficient internalization of neurophysiological theory might enable someone to be introspectively aware of her neurophysiological properties as neurophysiological properties. And if this is possible, it should certainly be possible for Mary the sensorily deprived neurophysiologist (or, I suppose, her ectoplasmologist sister) to imaginatively come to know, just in virtue of her scientific knowledge, what it's like to have color experiences. These, of course, would be the theory transformed experiences of the scientifically sophisticated, rather than of us scientific neophytes. But they would be introspectable experiences nonetheless.

Now I don't need to deny that sufficient training might enable one to become introspectively aware of her neurophysiological properties as such. But if my earlier criticism of Churchland and Churchland's 1981 opaqueness ploy is right, this awareness would not represent a more refined awareness of the same things we are now aware of as qualia. The possibility of such introspective awareness of neurophysiology thus does not ease the difficulties facing an identity theory.

IV

In the recent literature on this topic, there is a more or less standard way in which identity theorists try to explain the apparent contingency of mind-brain identity statements. I will call this the "standard anti-Kripkean strategy" (SAKS for short). According to SAKS, there is, corresponding to any mind-brain identity, an associated contingent identity statement with which the mind-body identity statement is confused. I will more or less follow Richard Boyd's (1980) account of how this associated contingent identity is constructed.

Suppose, to take the problem in its most general form, we want to explain the apparent contingency of (12). Let rs be the rfc for "having S(P)" and "(Ix)Fs(rs,x)" its rfd, and let ro be the rfc for "0" and "(Ix)Fo(ro,x)" its rfd. Then the contingent identity statement with which (12) is confused is:

\[(13) (Ix)Fs(rs,x) = (Ix)Fo(ro,x)\]

Of course, in order for this strategy to work, at least one of the rfd's must be non-rigid. In this section I will examine this strategy and some recent defenses against Kripkean arguments that perhaps use alternative strategies.

We have already encountered one unsuccessful example of SAKS in Section II. To propose that rs is contingently related to the referent still leaves us with a subjective rfc. But Richard Boyd (1980) has suggested that we might try a comparable move for the materialist (in the case of (12), the objectivist) designator. This might go as follows: Though the referent of "0" is an objective sentient property, its rfd involves reference to objective properties which are not sentient, and are thereby contingently related to the referent. Thus, "(Ix)Fo(ro,x)" is non-rigid, and (13) is contingent. (I do not, of course, want to say that
ro is objective and non-sentient, since that would both contradict (A) and beg the question against the conclusion I wish to establish).

This move assumes that the objectivist designator in (12) has an rfc which is only contingently related to its referent. But the whole point of Kripkean intuitions is that (12) seems to be contingent no matter what objectivist designator is used. Hence, if we replace any "O" which makes (12) true and has its reference fixed as Boyd suggests with a co-referential objectivist designator that does not have a merely contingently related rfc, the resulting identity statement would still seem to be contingent. But Boyd's explanation of this fact would then not work. So Boyd must assume that there can be no such co-referential objectivist designator. Consulting (D7), this means that, for any designator co-referential with the original "O", either (i) It does not have an objective property as referent (denial of (a) of (D7)), (ii) It is not known, or introduced in order, to designate an objective property (denial of (b) of (D7)), or (iii) Its rfc is contingently related to its referent. (i) is ruled out by the hypothesis that the designator is co-referential with the original one. And if any such co-referential designator, used by any cognizer, in any possible world falls into category (ii),¹² I don't see how we can fail to conclude that the referent is unmanifest. I think the same can be said about category (iii) as well, as the following considerations show. If an objective property, P, is manifest, then there is (for some cognizer in some possible world) an objectivist designator of it whose rfc is the property being-manifested-to-by-P. But being-manifested-to-by-P is necessarily related to P; i.e., it is essential to being-manifested-to-by-P that it is P that is manifested. Hence, if all an objective property's possible objectivist designators have rfc's merely contingently related to the referent ((iii) above), P cannot be manifest. So the success of Boyd's version of SAKS means that the identity theorist would have to be committed to the unmanifest nature of sentient properties. I suspect that this is a conclusion that many identity theorists would balk at in any case. (Consulting the definitions (D1)-(D6) shows what a very strong commitment it is.) More importantly, given (A) of Section I and the manifest nature of any rfc, it can't be true. There must be some sentient properties that are manifest—indeed, some that are manifest to human-type cognizers. So Boyd's version of SAKS won't work.

I fully realize that there is a step in the preceding argument that many will find contentious: namely, the step that says that being-manifested-to-by-P is essentially a manifestation of P. My answer is to simply stipulate this as part of the definition of being-manifested-to, and to stipulate that any manifest property has (for some possible cognizer in some possible world) a necessarily associated rfc for one of its designators. I think there are good reasons for such stipulation, but I won't go into them here. The important point is that, in the present context, this begs no important questions. What's controversial is whether or not there are any objective properties that are manifest in this sense.¹³ But my argument against Boyd nowhere assumes that there are; it only assumes that sentient properties are manifest. In the present context, which takes seriously the introspectibility which leads, e.g., to the inverted spectrum argument, I don't see how this can be faulted. Hence, the criticism of Boyd's version of SAKS goes through.¹⁴

These considerations spell trouble for an argument by Paul Churchland (1985a and 1985b) that may be intended (though it's not clear) as a further criticism of Kripkean arguments. It clearly is in-
tended as an additional criticism (to that already considered) of Jack-
son's 1982 argument. According to Churchland, Jackson's argument 
trades on an equivocation on "know about." Mary the sensorily deprived 
neurophysiologist may know all about neurophysiological properties, but 
not know all about qualia, simply because the first kind of knowing 
about is via a linguistic mode of representation whereas the second is 
non-linguistic. Hence, it does not follow that the neurophysiological 
properties and qualia are not identical.

Now I take it to be an implication of quale's being non-linguistic-
ally knowable that it is manifest. And I take it to be an implication of a 
neurophysiological property's, qua neurophysiological property, being 
known only linguistically that, even if the property is manifest, it is not 
in virtue of its being manifest that its neurophysiological designator has 
its reference fixed. Hence, neurophysiological designators have non-rigid 
rfd's. We can now easily see how one might try to use this distinction 
against Kripkean arguments. The result, in fact, would be the same ver-
sion of SAKS that Boyd employs, and the criticism of it would be the 
same. (In fact, the criticism would reach its culmination a bit earlier, 
because of Churchland's conceding that qualia are manifest.)

It may be noted that Boyd's strategy could save the day for ma-
terialism if we were to understand materialism in a weaker way than I 
am here—such as that corresponding to physical.14 Perhaps Boyd has 
something like this in mind, though I'm sure that Churchland does not. 
Nor, as I have remarked, do most materialists.

I conclude, then, that SAKS cannot explain the apparent contin-
gency of statements having the form (12). But let us now look at some 
recent criticisms of Kripkean arguments which, on first impressions 
(though perhaps not on second) seem not to rely on SAKS.

One such criticism comes from Christopher Hill (1984). Hill's proce-
dure is to first set up a series of plausible premises which yield the 
following conclusion:

(1) If there are no logical connections between the con-
cept of an F and the concept of a G, then it seems to 
be possible that being F ≠ being G (i.e., "being F ≠ 
being G" seems contingent)

Using the purported identity of pain and C-fiber stimulation as a repre-
sentative example, he then adds that the concepts of being a pain and 
being a C-fiber stimulation are not logically connected. From this it fol-
lows that the identity:

(I4) being a pain = being a C-fiber stimulation

is one that seems to be contingent. But this raises the question of 
whether or not the logical unconnectedness of the concepts expressed in 
(I4) is compatible with the truth of (I4). Now Hill's explanation of the 
logical unconnectedness of the concepts is given in the following pas-
sage:

(H1) . . . the concept of pain is formed in response to the 
experiences that one has in virtue of BEING IN a state
of mind that partially instantiates being a pain, and
the concept of C-fiber stimulation is formed in re-
sponse to the experiences one has in virtue of OBSERVING a brain state that partially instantiates being a case of C-fiber stimulation. (303)

This seems true enough, but it also seems to imply that the property of being a pain is subjective and that of being a C-fiber stimulation is objective. But Hill has more to say. If (14) should be true, then the situation described in (H1) amounts to the following:

\[(H2) \ldots \text{the concept of pain is formed in response to the experiences one has in virtue of BEING IN a brain state of a certain sort, and the concept of C-fiber stimulation is formed in response to the experiences one has in virtue of OBSERVING a brain state which is a state of the very same sort.} \quad (303)\]

Though the exact form of Hill's argument is not made too explicit here, as I reconstruct it he now in effect uses the following premise:

(2) If (H2) is possible, then the following are co-possible:

(a) (14) is (necessarily) true, and
(b) the concept of being a pain and that of being a C-fiber stimulation are not logically connected.

Combining this with (1) we get:

Therefore,

(3) If (H2) is possible, then the following are co-possible:

(a) (14) is (necessarily) true, and
(b) (14) seems contingent.

So if we can show that (H2) is possible, we have made our case.\textsuperscript{15}

Given the way Hill uses (H2), the following must hold: If we understand "concept of" opaquely, then the concepts of being a pain and being a C-fiber stimulation are concepts of different things; whereas if we understand "concept of" transparently they are concepts of the same thing. So (H2) implies a kind of first person/third person dualism of concepts for a single property that does not obtain for any non-physico-
logical property. E.g., my concept of being caucasian, having brown eyes, or weighing 150 lbs. is the same whether it's derived from my having those properties or somebody else's having them.\textsuperscript{16} If the case is different from brain states we need to know why. One possible explanation would be that the reference fixing for pain designators involves a subjective reference and no mediation by objective properties, whereas the reference fixing for brain property designators does involve mediation by objective properties. But this explanation both is incompatible with materialism and reduces Hill's strategy to SAKS. But Hill doesn't indicate how else we should understand his adherence to the possibility of (H2).

Terence Horgan (1984) also has a criticism of Kripkean arguments, and it bears some resemblance to Hill's. I won't go into the details of
his treatment, but just note that it involves two major problems. First, it deals with state identity rather than property identity, so it doesn't address the Kripkean arguments discussed in this paper. Second, and here is the resemblance to Hill—Horgan links "imaginative thinking" about qualia to a first person perspective and imaginative thinking about brain states to a third person perspective. As in the case of Hill, we need to be shown why this doesn't ultimately lead to the existence of states with subjective properties.

Now Horgan is sensitive to the fact that, on his treatment, property dualism is still a problem; though he doesn't arrive at this conclusion via a Kripkean argument, but rather via the considerations that originally led J.J.C. Smart (1959) to see it as a problem. But we can still see if his remarks on the subject throw any light on Kripkean arguments. His proposed solution is to maintain that the property of being a brain state and that of being a pain, for example, are the same, but that the associated predicates have different meanings. This would mean that the associated rigid designators have a common referent but different meanings. This can work, of course, only if the meaning is distinguished from the property designated. Horgan attempts to render this idea plausible in the following passage:

First, from the fact that some expressions have the same meaning and others differ in meaning, it does not follow that there are such entities as senses; perhaps there is only the relation of synonymy between expressions, and the positing of senses is gratuitous reification. Second, even if we do countenance senses, it is not obvious that they ought to be identified with properties; perhaps they should be identified with entities of a different sort, such as synonymy classes of expressions. (469)

But all this is singularly unhelpful. What we're looking for (whether or not it's Horgan's purpose to give it to us) is an explanation of how statements of the form (12) can seem contingent but be necessary. To explain this in terms of some irreducible and unexplained relations of sameness- and difference-in-meaning is to replace one mystery with another. We may just as well be told that there is an irreducible property of seeming-to-be-contingent. Furthermore, this doesn't even address the first person/third person dualism in "imaginative thinking" that Horgan concedes. Horgan's suggestion that meanings be identified with synonymy-classes is, if I understand it, equally unhelpful. He still seems to be taking synonymy as an unexplained and irreducible relation and constructing meaning on the basis of this. This still leaves one in the dark as to how an identity statement can seem contingent but be necessary (and again ignores the first person/third person dualism).

Perhaps Horgan would respond to these criticisms by claiming that, whatever meaning, or sameness- and difference-in-meaning are, we know this much about them: If two designators are different in meaning, then the identity statement involving them is, if true, knowable only empirically, and this is what seeming contingency amounts to. But, of course, the mere empirical knowability of necessary truths itself calls for explanation. SAKS has an explanation readily available—namely, that our knowledge of the truths in question is based on the contingently associated rfc's of the designators. I don't see that Horgan's rather
vague remarks offer an alternative (and, again, they don't address the first person/third person dualism issue.)

Despite their deficiencies, Hill's and Morgen's treatments do raise an interesting question. Is there some way to defend the apparent-contingency-though-actual-necessity of (12) by appealing to differences in meaning, concept, or the like, but do so without reducing this strategy to SAKS? But that's a line of inquiry I won't pursue, both for reasons of space, and because I think there are some more general considerations which cast doubt on the ability of any strategy apart from SAKS to accomplish what the identity theorist needs.

Any alternative to SAKS must be able to account for the apparent contingency of (12) without appealing to any contingent connections between rfc's and referent. So it must be able to account for such apparent contingency even when there is a necessary connection between rfc and referent for both designators. Now leaving aside the question of how referent and rfc might be necessarily connected (I have suggested one way beyond identity), I would assume that any such necessary connection must at least involve the following: that the rfc fixes the reference for a designator of that referent in all possible worlds in which that rfc occurs. That means that any alternative to SAKS must allow for the following scenario: Experiences e1 and e2 are necessarily reference fixing for two co-referential rigid designators. (More precisely, necessarily, if both experiences occur they are reference fixing for two co-referential rigid designators). But despite this, and, of course, the necessity of the co-referentiality, the designators don't seem to be co-referential. But this doesn't seem possible. If e1 and e2 are necessarily reference fixing for co-referential designators, one would suppose they must instantiate the same rfc, so that the necessity of the co-referentiality would be obvious.

Perhaps this is an insurmountable problem. Nevertheless, I wish to give the refutation of Kripkean arguments one more try.

V

Our discussion has in effect highlighted a first person/third person asymmetry in the way we know about qualia. This calls to mind an article written by Keith Gunderson (1970) which tries to reconcile first person/third person asymmetries with materialism. Put briefly, Gunderson proposes that we look upon the mind as a scanning mechanism. Now a scanning mechanism can include all sorts of things within its scope, including other scanning mechanisms. Thus, a given scanning mechanism, S1, can be an "object" to some other scanning mechanism, S2, and thereby appear on S4's "World List"—i.e., a list of descriptions of all the things about which S2 can acquire and store information via its scanning procedures. But S2 cannot fully include itself within its scope, and so would not include exhaustive information about itself on its World List. (Gunderson qualifies this, but the qualification isn't important for our purposes.) There is thus something comparable to a "first person/third person" asymmetry for such scanners, but this in no way tempts us to a "non-materialist" or otherwise "subjectivist" theory of scanners. Since Gunderson's proposal is that minds be understood as scanners, we should not be so tempted in our theory of mind either. Rather, "The difficulty in construing our self at any given moment as an
item wholly susceptible to third-person physicalistic and behavioristic
descriptions is comparable to the difficulty a periscope would face in at­
tempting to place itself between its own crosshairs". (300) Such a dif­
ficulty reflects a logical impossibility and has no metaphysical implica­
tions.

Let’s see if we can apply this idea to the problem of subjectivity
as it has emerged in this paper. There are some difficulties in doing so.
First, Gunderson more or less interchangeably talks about at least the
following three kinds of subjectivity: the kind associated with the so­
called "systematic elusiveness of the I" (the "ego-elusive" kind), that
associated with de se attitudes, and the kind which has been the topic
of this paper. Maybe these are all connected in some way, but if so,
Gunderson doesn’t show how. So even though it’s clear how his proposal
accounts for ego-elusive subjectivity within a materialist framework, it’s
not clear (despite some interesting and suggestive remarks on his part)
how it’s supposed to work in the other two cases. This is especially
true for the kind of subjectivity under consideration in this paper.
After all, qualia are "scannable" from the first person perspective, and
this would appear to make Gunderson’s approach irrelevant.

This brings up a second difficulty. I have attributed
Gunderson a belief in "ego-elusive" subjectivity, but he is really somewhat ambigu­
ous as to what it is that is elusive from a first person perspective. Sometimes he talks as if it’s the "ego" (i.e., mind) and sometimes as if
it’s certain states and features of the ego, indeed, even qualia them­
selves. But to concede the existence of qualia is to concede the exist­
ence of features which are scannable from a first person perspective. So
Gunderson’s approach seems thereby to be even further off target.

But maybe something helpful can be salvaged from all this. I will
ignore Gunderson’s apparent remarks about the first person elusiveness
of qualia, and consider only ego-elusiveness, and I will sketch a possi­
ble connection between this and the apparent subjectivity of qualia
which has emerged in this paper. This connection, if it can be worked
out, holds out some promise of giving the identity theorist what he
needs to refute Kripkean arguments. But some problems will have to be
overcome if this promise is to actually be fulfilled.

Recall from the discussion of Section II that the manifestation of a
property is always a manifestation qua being instantiated by some indi­
vidual. Usually this means that the manifestation of the property (to a
cognizer) goes hand-in-hand with the manifestation of the individual
which instantiates it. Call this "individual-bound" manifestation of a
property. But in the discussion in Section II I also left it open that a
property could be manifested to a cognizer without the individual which
instantiates it being so manifested. I now want to exploit this idea. I will
call such manifestation "individual-free" manifestation.

If Gunderson is right, that which instantiates qualia (the mind,
the ego, or whatever) cannot be manifested to itself. So any manifesta­
tion of qualia to the individual who has them must, logically, be individ­
ual-free. Suppose further that individual-free manifestation of properties
can, logically, occur only in such first person cases. Then if manifesta­
tion of others' qualia to an individual is possible, it must, logically, be
individual-bound. There is thus a logically necessary first person/third
person asymmetry in the manifestation of qualia. But this asymmetry no
more has untoward (for the materialist) metaphysical consequences than does the ego-elusiveness upon which it is based.

We can see how the materialist might exploit this idea to explain the apparent-contingency-though-actual-necessity of (12). The idea would be that the rfc's for the designators are radically different manifestations reflecting a logically necessary first person/third person asymmetry. Perhaps this is what Hill and Horgan (and Nagel—see n. 15) are trying to capture in linking the materialist answer to Kripkean arguments to the distinction between being in a psychological state and observing that state. And since individual-free manifestations apply only to properties of the mind, we have an explanation of why the consequent first person/third person dualism of concepts holds only for the psychological properties.

All this needs to be made intelligible, of course. For one thing, it needs to be shown that the idea of individual-free manifestation of properties really does make sense. For another, it needs to be shown that a materialist can non-arbitrarily distinguish between a unified, individual mind which has psychological properties, and the rest of the body that doesn't. Even if all this can be done, moreover, it's not clear that a materialist answer to Kripkean arguments is forthcoming. The proposed strategy, as much as any other, must contend with the general objection to non-SAKS materialist strategies offered at the end of the preceding section. The answer that must be defended is that individual-bound and individual-free manifestation can necessarily be manifestations of the same property, even though they themselves don't have any property in common. Intuitively, I am not unsympathetic to such a claim, but it has to be shown how this can be so, and I am not presently capable of doing so. Until it is done, I must conclude at least that no way has so far been found of overcoming Kripkean arguments against identity theories. If no such way should be forthcoming, we are left with three possible ways of accounting for qualia: make good on a causal/functional account of them; explain them away; or regard them as constituents of the irreducibly subjective half of a property dualism.

ENDNOTES

1 An earlier version of portions of this paper was read at the 1985 Western Division Meetings of the American Philosophical Association. Rod Bertolet was commentator.

2 Despite his adoption of the same terminology, Herbert Feigl's (1958) "physical" may not be exactly the same as Meehl's and Sellars'. E.g., Feigl says that something is physical "... which can be described (and possibly explained or predicted) in the concepts of a language with an intersubjective observation basis" (1958, 421) and this seems to be stronger than what Meehl and Sellars say. In this paper I will stick with Meehl's and Sellars' definition.

3 Another such way is proposed in Holman (1986).

4 Notice that even though all rfc's are manifest, I have not said that all manifest properties are rfc's. Hence, I leave it open that there may be manifest properties that are not features of experience and
hence are directly observable but not introspectible. See again Holman (1986).

5 For another demonstration of the futility of maintaining a contingent relation between the rfc and referent of the sensation designator, see n. 14 and the attendant text.

6 Or, more correctly, I think it is so manifest qua apparently being instantiated by some individual. E.g., I would wish to leave open the possibility that colors are being manifested in visual experiences even if the proper interpretation of modern physics is that it tells us that colors are not instantiated.

7 Possibly something even slightly stronger is at stake, since it does not obviously follow from the definitions that a subjective property must actually be instantiated by a mind.

8 And see n. 11 of Nagel (1974), where the author also suggests that subjectivity is the real point of Kripkean arguments.

9 In Nagel (1965) and perhaps in some of his remarks in Nagel (1979), Nagel addresses the kind of subjectivity associated with what David Lewis (1979) calls "irreducibly de se attitudes." This is clearly a stronger sense of subjectivity than the one I have in mind. Frank Jackson in Jackson (1982) claims to detect the de se sense of subjectivity in Nagel (1974), but I don't, and in any case I think Nagel (1974) at least also addresses a kind of subjectivity much closer to my own.

10 Jackson, in Jackson (1982), rightly takes Nagel to task for sometimes being sloppy—or just plain wrong—in his formulation of this point. Nagel sometimes talks as if the question is whether or not one could imagine what it's like to be a bat without having had bat-like experiences. But that's not the point, or shouldn't be. After all, one might have an innate capacity to imagine, or extrapolate to imagining, bat-like experiences despite not being able to have them. But hereafter I will interpret Nagel as I have in the text above.

11 This point is made explicitly in Nagel (1979). It is also made in Nagel (1965) but, as we have seen, here Nagel is talking about a different kind of subjectivity.

12 Of course, given the falsity of (i), (ii) probably entails (iii).

13 The claim that there are invokes the old idea of a perceptual experience's "phenomenological content" in such a way as to challenge some currently popular views about the reference fixing role of perceptual experiences via a via objectivist designators. These popular views construe perceptual experience as reference fixing for designators of their typical causes. (Contrast this with my remark in n. 6) For a fuller discussion, see Holman (1986).

14 A parallel argument could have also been used to show the futility of the version of SAKS considered earlier, which appeals to the non-rigid nature of "((x)Fs(rs, x))."

15 Cf. Grover Maxwell (1978) where the author adopts Boyd's strategy but concedes that it does not yield a full-fledged materialism. Maxwell
EMMETT L. HOLMAN

calls his theory "non-materialist physicalism", where "physical" does indeed mean "physical".

16 Nagel advocates a way of explaining away the apparent contingency of mind-brain identities which bears some resemblance to Hill's, and to Horgan's which will be discussed shortly. Or rather, he thinks that his approach works provided we have already explained how such identities can be true. See n. 11 of Nagel (1974). I don't follow Nagel here. As I see it, Kripkean arguments highlight the mysterious nature of mind-brain identities by focusing on their apparent contingency. If the contingency can be explained away, so can the mystery. Another reason for my puzzlement with Nagel is given in n. 17.

17 Nagel seems to recognize this point in a parenthetical remark in n. 11 of Nagel (1974). This makes it even more puzzling that he endorses something like Hill's strategy.

REFERENCES


Churchland, P.M. and Churchland, P.S., 1981, "Functionalism, Qualia and Intentionality", Philosophical Topics 12, 121-45.


