

Max Black's Objection to the Identity Theory of Mind

By Ronald Planer, Rutgers University

In "Sensations and Brain Processes,"¹ J.J.C. Smart defends the identity theory of mind (hereafter, the identity theory) from a series of dualist objections. The objection Smart considers to be the strongest he attributes to Max Black. Roughly, Black's challenge to the identity theory is how to account for the epistemic gap, that is, how we can know a lot about mental states without knowing anything at all about neuroscientific states and vice versa if psycho-physical identities are true; or put somewhat differently, why it is that statements of the form 'M = B' (where 'M' refers to a mental state and 'B' refers to a brain state) are, if true, knowable only *a posteriori*.²

At the time Smart was writing his paper, the standard explanation for the "appearance of contingency" of *a posteriori* identity statements was that they were cases in which one and the same thing was described in terms of two logically distinct sets of properties. For example, on this way of thinking, the reason that the statement 'lightning = electrical discharge' is knowable only *a posteriori* is that we use two logically independent sets of properties to pick out the referent of these terms, its lightning properties (being a bright flash of white light), and its electrical discharge properties (the properties our scientific theory of electricity bestow upon it). Thus, Black takes his objection to show that even if it's true that there are no mental substances (a main tenant of the identity theory) there must still be mental *properties*. Otherwise, identity statements between mental and neuroscientific states would not seem contingent.

This paper considers Black's objection and also looks at Smart's reply to it. In order to successfully meet the objection, Smart must find a way to either "explain away" this epistemic gap as apparent or find a way to accept it without allowing that there are mentally irreducible properties. Smart opts for the latter strategy, claiming that talk of the mental is "topic-neutral."³ I argue that although his strategy provides us with one way of responding to Black, it is untenable on other grounds. My reasons for thinking this boil down to two. First, it seems that the only essential property of sensations is that inner qualitative feel to the state, that special what-it-is-like-for-a-subject aspect, yet Smart's topic-neutral analysis ignores this feature of sensations altogether. Second, it again seems intuitively obvious that we are not guided by any topic-neutral description in picking out or sensations but are rather guided only by the very *feel* of the state. This suggests to me that Smart's account is bound to fail.

The structure of this paper, then, is this. First, I explain what exactly the identity theory holds. Second, I present Black's objection to it. After that, I turn to Smart's reply to Black. And finally, I assess the strengths and weaknesses of Smart's reply. Importantly, I intend for my criticisms to be limited to Smart's treatment of sensations as opposed to other mental states like beliefs and desires. For as I see it, the topic-neutral approach may correctly analyze propositional attitudes, but it cannot give an adequate

account of sensations such as pain.

I. The Identity Theory⁴

According to the identity theory, states and processes of the mind are identical to states and processes of the brain. Take the experience of being in pain, for example. The claim is *not* that this experience is correlated with some brain process. For a correlation would imply that there are two things, namely, the experience of pain *and* a brain process, which is false if the identity theory is true. Rather, the claim is that the experience of pain (or the feel of pain) *just is* a brain process. Or another way to make the same point is to say that inasmuch as a mental term such as 'the feel of pain' or 'the sight of a sunset' refers to a process which goes on in us under certain conditions, the process it picks out *just is* a brain process. The sense of identity here is to be understood in the strict or numerical sense, as opposed to, say, mere spatio-temporal continuity. (An example of this latter sense of identity might be the relation which holds between a statue and the portion of clay it's made of.) As such, were we to count all the brain events in the world, no mental event would be left uncounted (though not every brain event would be a mental event).

In addition, the claim is *not* that talk about the mental can be translated into talk about the brain. Actually, the identity theory holds just the opposite: i.e., it holds that (i) mental terms do not *mean* the same thing as brain terms (e.g., 'the feel of pain' does not mean the same thing as 'C-fiber stimulation'), from which it follows that (ii) sentences containing mental terms cannot be translated into truth-functionally equivalent sentences containing only brain terms. In this way, the identity theory differs from both analytical behaviorism and analytical functionalism. (Not that these theories claim that talk about the mental can be translated into talk about the brain; they don't. However, they are theories about the *meaning* of our mental terms, which Smart says the identity theory is not.) Importantly, this implies that no objection revolving around the idea that reports of sensations are not reports of brain states or processes can refute the identity theory.

This claim should not cause us immediate concern. For example, as Smart points out 'the Morning Star' and 'the Evening star' do not mean the same thing, and yet the Morning Star and the Evening Star are identical. Or, alternatively, 'lightning' and 'electrical discharge' do not mean the same thing, but we now know that lightning just is electrical discharge. Accordingly, we should not expect that sentences containing one of the terms in these pairs can be translated into truth-functionally equivalent sentences containing only the other term in the pair. Put somewhat differently, we know that sentences containing the term 'river X' cannot be translated into sentences containing only the term 'Y number of H₂O molecules,' yet this alone does not show that river X is something over and above Y number of H₂O molecules.

However, these examples do show that Smart rejects a direct reference theory of meaning, i.e., one on which the meaning of a term is exhausted by the referent of that term, in favor of a descriptivist theory of

meaning, i.e., one on which a term reaches its referent in virtue of some descriptive content (e.g., a Fregean sense, a definite description) associated with the term. To bring out the difference between the two views, consider the following. On the direct reference theory, the name 'the Morning Star' just means the planet Venus, whereas on the descriptivist theory, this same name might mean something like 'the last star to disappear in the morning,' and it is in virtue of this descriptive content associated with the name that reference to Venus is secured. The important point is that this move allows Smart to claim that even if sensation terms such as 'the feel of pain' and brain terms such as 'C-fiber stimulation' do not do not *mean* the same thing, they may still *refer* to the same process. It follows that the identity theory cannot be ruled out on purely *a priori* grounds (for *a priori* reasoning deals only with the intensions of our terms.)

II. Max Black's Objection (Objection 3) to the Identity Theory

The first two objections that Smart considers are intimately related. Together, they establish that (a) if true, identity statements between mental terms and brain terms are knowable only *a posteriori*, and (b) mental terms and brain terms do not mean the same thing.⁵ (Basically, (b) follows from (a).) Presumably, (a) is the case because we describe some process P *qua* mental process in a different way (that is, by appeal to different properties) than we describe P *qua* brain process. The identity statement 'M = B' (where 'M' refers to a brain state or process and 'B' refers to a brain state or process) is knowable only *a posteriori*, then, because the properties which we use to pick out P *qua* mental process are distinct from the properties which we use to pick out P *qua* brain process, and it could have been the case (in the epistemic sense of 'could') that these properties were in fact instantiated by distinct processes; particularly, the properties which we use to pick out instances of M could have been instantiated by something other than a brain process, epistemically speaking. And we have already said something about (b): the fact that we can imagine or conceive of instances of M without instances of B and vice versa, merely shows that 'M' and 'B' do not *mean* the same thing. But Smart makes no claim that mental terms and brain terms mean the same thing, only that they refer to the same states or processes.

Black's objection, then, is as follows. Suppose the psycho-physical identity statement 'pain = C-fiber stimulation' is true. Then we know that 'pain' and 'C-fiber stimulation' refer to the same process P. The truth of this identity statement would then imply that the event of being in pain is nothing over and above the state of undergoing C-fiber stimulation, which is, of course, a physical process in the brain. This is the desired result of proponents of the identity theory. But from (a) and (b) above it follows that the properties which we use to pick out instances of P *qua* pain are distinct from the properties which we use to pick out instances of P *qua* C-fiber stimulation, and if Smart now wishes to reject either (a) or (b), the identity theory is refuted; it can't respond to the first two objections of the paper. Or, put differently, if we think as Frege did that the meaning of a term reflects the way in which the referent of that term is presented to us in experience

(what Frege called the "mode of presentation"⁶), then it follows that (a) and (b) are explained by P presenting itself to us one way *qua* pain and another way *qua* C-fiber stimulation. The problem for the identity theory is this: Doesn't that leave us with a set of mentally irreducible properties in virtue of which P is presented to us as pain?⁷

We can put Black's challenge to Smart more precisely still: Smart admits that (i) 'pain' and 'C-fiber stimulation' are coreferential, and thus (ii) there is one process P which is picked out by both of these terms, but Smart also says (iii) these terms do not mean the same thing, and thus (iv) this difference in meaning (which accounts for the *a posteriori* of psychophysical identity statements) reflects the different ways (the different properties by virtue of which) P presents itself to us in experience. Suppose then that P presents itself to us *qua* pain by virtue of one set of properties F while it presents itself to us *qua* C-fiber stimulation by virtue of a different set of properties F'. If 'pain = C-fiber stimulation' is true, then it follows that the pain is a physical process, and thus one half of the identity theorist goal is met. However, if even one of the properties in the set F (i.e., the set of properties involved in P's presentation as pain) are mentally irreducible, then the identity theory has failed to avoid dualism; it must admit that there are non-physical properties of some brain states and processes, namely, those that are also mental states and processes.

In short, then, the challenge Black poses to the identity theory is to explain how the way P presents itself to us *qua* pain does not involve any properties which are mentally irreducible and which therefore cannot be covered by the identity theorist's ontology. To be sure, this is a tough task for among other things something about the set of properties F explains why it is that one can have seemingly complete knowledge of the mental without knowing anything at all about the brain. How does Smart respond to Black?

III. Smart's Reply to Black⁸

Smart claims that our concept of pain (roughly, the way we have of conceiving of pain as being) is *topic-neutral*, i.e., neutral between a dualistic and physicalistic metaphysics. That's to say, our concept of pain does not commit us to pain having only physical properties, nor does it commit us to pain having any mentally irreducible properties. The same goes for all other sensations, and more generally, mental states such as beliefs and desires.

For example, Smart says that when we report a sensation such as pain, what we are really reporting is something along the following lines: that there is something going on in me which is caused by my body sustaining physical damage, which is apt to make me to shout "Ouch!", and to avoid the thing in my environment which is the cause of the damage. This formulation leaves open the question as to whether the process I report is physical or non-physical. (The 'something' as it occurs in this phrase is like an abstract place-holder for a yet to be identified process.) Smart claims that it is simply more plausible given the findings of contemporary brain science to suppose that this topic-neutral description picks out a physical process as opposed to a non-physical process. Our concept of pain, then, amounts to nothing more

than a very abstract description of *something* or *some process* which goes on in us under certain conditions C and is apt to cause effects E.

Notice how this account explains why psycho-physical identity statements are knowable only *a posteriori* (i.e., (a) above) and also why mental terms and brain terms are not equivalent in meaning (i.e., (b) above). Consider (b) first. P, that is, the process picked out by 'pain' and 'C-fiber stimulation' does indeed present itself to us in different ways (in virtue of different properties). *Qua* C-fiber stimulation, P is presented to us in terms of its neuroscientific properties such as having chemical composition X and structure Y, whereas *qua* pain, P is presented only very abstractly to us, as *something* which goes on in us satisfying the following list of properties . . . (where '. . .' is a list of functional properties such as being caused by bodily damage, being apt to make one shout "Ouch!", and so on). This explains why 'pain' and 'C-fiber stimulation' do not mean the same thing.

Also, now that we've seen why mental terms and brain terms differ in meaning, we also get an explanation of (a), i.e., since *qua* pain we pick out P by virtue of a set of properties F, while *qua* C-fiber stimulation we pick out P by virtue of a different set of properties F', we have to discover via empirical means that one and the same process instantiates both sets of properties. Analogous remarks apply to cases like 'the Morning Star = the Evening Star': we had to discover that the property of being the last star to disappear in the morning is had by one and the same celestial body which has the property of being the first star to appear in the evening. No amount of *a priori* reasoning could tell us this since there is nothing in the meaning (intension) of these terms to suggest that they are identical. Of course, this also explains why psycho-physical identity statements would appear to be contingent even if they were true necessarily. They appear contingent because the properties which we use to pick out pain could (in the epistemic sense of 'could') have been instantiated by some process other than C-fiber stimulation, just as the properties which we use to pick out the Morning star could (in the epistemic sense of 'could') have been instantiated by some celestial body other than the Evening Star.

IV. Evaluation of Smart's Reply

Thus it would seem that Smart's response succeeds in one jump. But are there really no problems with his topic-neutral analysis of the mental? In this section, I consider an objection to Smart and the various ways he might respond to it.

Objection: It is unclear to me whether our concept of pain even involves the functional properties that Smart cites in his topic-neutral analysis of this concept. For example, if they were part of our concept of pain, then we would expect a sentence such as "Jones is in pain" to entail the sentence "Jones has something going on in him which causes him to shout 'Ouch!'" which I think is implausible. Or, to put it differently, we'd expect a sentence such as "If Jones is in pain, then he will avoid the thing causing damage to his body," to be true *a priori*, which again, I also think is implausible. (Notice

that Smart can no longer say that the identity theory is not a theory about the meaning of our mental terms. That's to say, in purporting to give an analysis of what we mean by reports such as "I am in pain," Smart has explicitly committed himself to a semantic thesis about our mental terms and concepts.) But putting these worries to the side, I think there is a more serious problem with Smart's topic-neutral analysis of pain. The problem is this. Even supposing that the functional properties which Smart uses to analyze our concept of pain are in fact part of our concept of pain, it is clear that these are not the properties by virtue of which we pick out our pains. Rather, the only property which appears to guide us in picking out instances of our pain is *that feeling* of pain, the feeling we can focus our attention on when we are actually in pain. Yet the topic-neutral analysis ignores this feature of our pain experiences altogether. In short, it seems to me that what Smart has done is to describe pain by its accidental properties alone, leaving the only property which is essential to it, namely, that feeling, out of his analysis. And it is precisely the property of being that feeling of our pain experiences which is the most likely to be non-physical.

Roughly, there are two lines of response Smart could take to this objection: either (1) deny that that feeling is a real property of our pain experiences, or (2) claim that a topic-neutral analysis of pain can include this property in addition to the other functional properties. Let us consider each of these possibilities respectively.

Smart might reply to this objection by saying that we are making some kind of mistake if we say that that feeling is a real property of our pain experiences, a 'real property' in the sense that being a triangle is a real property of some geometric figures. However such a response is at best highly implausible. For if I ask you to pinch your arm and then focus your attention on the feeling of that experience, you will certainly have no trouble understanding what I am asking you to do. Furthermore, it is doubtful that you would be willing to concede that you are "mistaken" about the occurrence of that feeling, at least in the absence of any other assumptions. At any rate, surely the burden is on he or she who wishes to flatly deny that this property is real, that it is an illusion of some kind. For, all other things being equal, it seems that if we can know anything for sure, we can know that each of our sensations has a certain intrinsic, qualitative character to it, that is, that there is something-it-is-like to have those experiences which goes beyond the mere functional properties which the topic-neutral approach seeks to analyze such experiences with.

To shift this burden, such a hard-nosed skeptic could ask us to flesh out what we mean by 'that feeling'. Unfortunately for the common-sense view, it is difficult to comply with the skeptic on this point. However, one response would be as follows. First, it appears that our concept of pain and also of other sensations is primitive, i.e., it is not built up out of any lower-level concepts into which it can be analyzed. This would explain, for example, why it is that we cannot describe that feeling as being like anything else in the world, since in order to explain how A is like B, there must be some part of A which is like B. Of course, it seems that we can describe that feeling of

pain as like that feeling of an itch or a tickle, though more intense. But this would obviously not please such a skeptic, for our account would then turn on an unexplained explainer. However, since our criteria for the application of the concept 'that feeling of pain' is one based purely on introspective data, it seems unfair to ask us to elaborate on the details of this concept using only terms or concepts which have a public criteria for application. So I am not sure that proponents of the common-sense view actually have any dialectic obligation to the skeptic on this point.

I am not sure how the skeptic could respond to these considerations. Perhaps we have reached a point in the debate where there is a standoff of intuitions. But I am relatively sure that, if pushed, Smart would not want to deny the existence of such qualitative properties of our sensations anyway. For, after all, his aim is to *explain* the mental, not to *explain it away*. Therefore, let us turn to possibility (2), the idea that properties such as being that feeling of pain can be accounted for within the topic-neutral framework.

Before we said that Smart claims that talk of the mental is always topic-neutral, i.e., neutral between a dualistic and a physicalistic metaphysics. This is because it conceives of a given mental state or process M as merely some state or other which plays such-and-such a functional or causal role in our behavior. The functional conditions that a state must fulfill to count as M of course vary depending on the mental state in question. Above we gave an analysis of pain, but similar analyses could have also been given for other sensations. Sticking with pain, then, the above objection argued that there is at least one property of our pain experiences, namely, that feeling, which Smart's topic-neutral analysis of leaves out. However, it's not clear that Smart's account can't be patched up in some way so as to answer our objection and other objections like it. For example, couldn't we just say that when I report that I am in pain, what I am really saying is that there is something going on in me, whatever it is, which instantiates the property of being caused by bodily damage, being apt to make me shout "Ouch!", being apt to make me avoid the think in my environment which is causing that damage, and being that feeling? That's to say, is there any a priori reason which speaks against including the property of being that feeling of pain in a topic-neutral analysis of pain?

Several worries rise to the surface here. This might get the semantics of our concept of pain right, but remember that Smart wants to claim that the 'something' in the topic-neutral analyses actually picks out physical states or processes of the brain. Yet, as the above objection argued, the property of being that feeling is the most likely to be a non-physical property. Why think this? It seems that that feeling is likely to be non-physical mainly because no amount of neurophysiology could ever reveal the state in you which has this property. That's to say, we'd never be able to open up the brain and see that such-and-such a brain state or process instantiates the property of being 'that feeling.' Hence, at first blush, it would appear that that feeling is not located anywhere in physical space (which is a necessary condition for all physical properties, I assume). Moreover, on this account, it would just have to be a brute fact about some physical states and processes that they feel a certain way, which is obviously tendentious.

However, if we don't leave that feeling simple and unanalyzed in our analysis of pain, it seems that we will always be able to ask: Even if all the functional criteria for pain are met by some process P, does that process really feel like our pain feels? Might it not meet all these conditions and yet be completely devoid of qualitative character? I think these are perfectly intelligible questions, and the fact that they are intelligible raises worries for any topic-neutral analysis of a sensation that does not leave whole and unanalyzed the property *being that characteristic feeling of the sensation*. (For example, if the sensation we sought to analyze was an itch, then the relevant property would be being that feeling of an itch, and so on.)

In short, I do not think that Black's objection to Smart succeeds in showing that there *must* be properties of our mental states and processes which are irreducibly mental. I think Smart is right insofar as he claims that his topic-neutral analysis of sensations and other mental states explains why both (a) identity statements between mental terms and brain terms are, if true, knowable only *a posteriori*, and (b) why mental terms and brain terms do not mean the same thing. That's to say, if we grant Smart the premises of his argument, we get an adequate answer to (a), i.e., since our concepts of sensations such as pain are so different from our concepts of brain processes, there is no way to *a priori* deduce from these concepts that sensations are identical to brain processes, even if they in fact are, and we also get an adequate answer to (b), i.e., the meaning of mental terms and brain terms reflect the different ways mental phenomena and neuroscientific phenomena are presented to us in experience, the prior only very abstractly, the latter in concrete, objective-scientific terms. However, it is unclear to me whether Smart's account of the mental succeeds in removing all doubt about there being mentally irreducible properties such as that feeling of pain. Smart needs more of a story here.

V. Conclusion

The purpose of this paper was two-fold. First, I wanted to explore Black's objection to the identity theory and Smart's reply to it. Insofar as Smart puts the objection in "Sensations and Brain Processes," I am convinced that Smart's reply to Black is at least satisfactory. I agree with Smart in that, just because identity statements between mental terms and brain terms are, if true, knowable only *a posteriori* and just because mental terms and brain terms do not mean the same thing, this alone does not show that there must be mentally irreducible properties. It does show, however, that our way of conceiving of mental states and processes is different from our way of conceiving of states and processes of the brain. But Smart has an explanation of this which does not concede that there are non-physical properties.

Second, I wanted to consider Smart's positive account of the mental. In response to his conceptual analysis of sensations, I developed some criticisms which I think show that Smart's topic-neutral approach is bound to be inadequate on independent grounds. Specifically, I objected over the fact that his analysis of, e.g., pain, leaves out the intrinsic, qualitative feel of the

sensation. Whether or not Smart's account can be patched up accordingly so as to meet these objections remains to be seen.□

© Ronald Planer 2008

[Return to the Table of Contents](#)

¹ Smart, J. (1959). "Sensations and Brain Processes." In D. Chalmers, editor, *Philosophy of Mind: Contemporary and Classical Readings*, New York: Oxford University Press, pp. 60-68.

² Ibid., p. 63.

³ Ibid., p. 64.

⁴ Ibid., pp. 61-62, for a discussion of the main points in this section.

⁵ Ibid., pp. 62-63.

⁶ Frege, G. (1892). "On Sense and Nominatum." In A. Martinich, editor, *The Philosophy of Language*, New York: Oxford University Press, Fifth Edition, p. 218.

⁷ See note # 2.

⁸ See note # 3.