MIND-BRAIN IDENTITY AND THE NATURE OF STATES

David Hunter

According to what I will call the ‘Identity Thesis’, a human’s psychological states are identical with states of or in her brain. Psychological properties, such as being in pain, believing that water is wet, or desiring ice cream are, on this view, neurological properties. Proponents of this view, such as David Lewis and David Braddon-Mitchell and Frank Jackson, admit that the Identity Thesis is costly, since it entails that beings with relevantly different nervous systems than ours, or without such systems at all, cannot share our psychological states, no matter how sophisticated their behaviour might be. But, they insist, this cost is worth paying. For, they contend, accepting it is the only way to guarantee that human psychological properties are causally efficacious.

I believe that this analysis of the costs and benefits is mistaken. I will argue, first, that the costs are higher than admitted since the view implausibly entails that the subjects of psychological states are neurones or brains. I will then argue that the advertised benefit is not delivered. There is, I will argue, little reason to anticipate discovering a neurological property of the kind advertised. Seeing all of this will require getting a bit clearer on what a state is, and on what it is to affirm identity in the case of states. For the initial allure of their view rests, I think, on various confusions about the nature of states.

I. The Logic and Ontology of States.

Talk of states is just a variation on talk of the instantiation of a property (or relation) (at a time).¹

To say that a thing is in a certain state is just to say that it has a certain property. To say, for instance, that Jones is in a state of believing that water is wet is just to say that Jones has the property of believing that water is wet, or that he believes that water is wet. And to say that Jones has the property of believing that water is wet is just to say that he is in a state of believing that water is wet. Following [Robinson 1990], I will adopt the following linguistic convention for talking about states.

An expression of the form ‘<s, F>’ will indicate that state (if it obtains) of the subject s, which is its (state of) being F.

This can be read as either: ‘S’s state of being F’, or ‘S’s being F’, or ‘S’s F-ness’. This expression refers to the state. It denotes it, if you like. This convention allows us to talk

¹ I will ignore relations and times in what follows. I will also assume that such phrases as ‘believes that water is wet’ express properties. If one is more comfortable taking them to express relations to propositions, then read my uses of ‘believes that water is wet’ in what follows as expressing the (relational) property of being related by belief to the proposition that water is wet. My discussion is indebted to that in [Robinson 1990], and to the account of events in [Kim 1993].
about particular states, such as Jones’ believing that water is wet. These are sometimes
called state tokens. A state token is an instantiation of some property in some thing. But
we also talk about state types. We might, for instance, say that Jones and Daniels are in
the same psychological state. We don’t mean that Jones’ believing that water is wet is the
very same particular state as Daniels’ believing that water is wet, just that those particular
states are of the same type. For this, we can introduce the schema ‘〈x, F〉’ where ‘x’ is a
free variable. Replacing ‘F’ with a predicate yields an expression indicating a state type.
For instance, ‘〈x, is wet〉’ indicates the state type of being wet, and ‘〈x, believing that
water is wet〉’ indicates the state type of believing that water is wet. These are state types
that many subjects may be in. In effect, as this convention indicates, talk of state types is
just inelegant talk of properties.

Talk of state tokens is closely related, logically, to talk of facts. For the following is a
plausible thesis.

For all s and F, 〈s, F〉 obtains just in case it is a fact that s is F.

Are states, then, just facts? Does a referring phrase of the form ‘S’s’s being F’ simply refer
to the fact that S is F? One might hesitate since a phrase of the form ‘S’s’s being F’ or ‘S’s
F-ness’ could not everywhere substitute for the corresponding sentence ‘S is F’. For
instance, although ‘Adams is surprised (knows) that Jones is wet’ is grammatically fine,
‘Adams is surprised (knows) that Jones’ wetness.’ is not. But this is merely grammatical.
For replacing ‘that’ with ‘at’, ‘by’ or ‘of’ yields the grammatically correct sentences:
‘Adams is surprised at (by) (knows of) Jones’ wetness.’.2 If one thing surprised Adams,
then the fact that surprised him is also the state that surprised him. The linguistic
difference between ‘S is F’ and ‘S’s’s being F’ is that while the former (if true) states the
fact that S is F, the latter refers to, or denotes, that fact.

The Identity Thesis holds that a person’s psychological states are identical with brain
states. What is involved, logically, in such a claim of identity? What does it mean to say
that state A is identical with state B? In the case where 〈x, P〉 obtains and 〈y, Q〉 obtains,
the following is, I think, a plausible principle about state identity.

〈x, P〉 = 〈y, Q〉 if and only if x = y and P = Q.

According to this principle state A is identical with state B, only if the subject of A is
the subject of B, and the property instantiated in A is the property instantiated in B. I think
this principle has considerable initial plausibility. If, as I have suggested, a state just is a
property instantiation, then for anything to be a given state it must be the instantiation of
the same property by the same subject (at the same time).3 The Principle says that the
property instantiated in state A must be identical with the property instantiated in state B if

2 Compare: ‘Adams was surprised that Clinton was impeached’ and ‘Adams was surprised by
Clinton’s impeachment’; ‘Frege was shocked at Logicism’s failure’ and ‘Frege was shocked that
Logicism failed.’

3 Many important questions about the nature of states remain to be answered. A complete account of
the nature of states would explain, inter alia, what is essential to a state and whether events are
states. But for present purposes these questions need not be answered. The general issue of the
nature of states is insightfully addressed in [Steward 1997], Chapter 4.
A and B are identical. I will assume that property A is property B just in case it is impossible for a thing to have one of them and yet for it not to have the other. At the very least, being necessarily co-instantiated is a necessary condition on the identity of properties. This fact about property identity is, of course, of considerable importance in the context of the Identity Thesis.

II. The Identity Thesis

According to David Braddon-Mitchell and Frank Jackson, ‘your desire for ice-cream, your pangs of hunger, and your believing that the lights have turned green are all states of your brain.’ [Braddon Mitchell & Jackson, 1996, p. 91] David Lewis asserts that, on his view, ‘mental states are contingently identical to physical, in particular, neural—states.’ [Lewis, 1994, p. 412] They arrive at this view as follows. Psychological states are, they claim, essentially states that play a certain causal role, states that were caused to obtain in some specific way, and are apt, together with other states and events, to bring about still other states and events. The states that play such roles are, they contend, at least in humans, states of the brain or of the central nervous system. And so, they conclude, psychological states are (at least in humans) states of the brain or of the central nervous system.4

Let me make this more concrete. Consider Jones’ state of believing that water is wet. According to this view, the following states obtain. Inside Jones, there is some part or other that has some property or other. Since that part is standardly assumed to be a neuron or group of neurones, I will call it ‘n’.

\(<n, P>\)

P is a neurological property, a property of a neuron or neural cluster. This state is called the realiser state. It is the state that plays the relevant causal role. It was caused to obtain in a certain way (or in one of several ways) and it (or its subject) is apt, in combination with other internal states, to have certain effects. Calling this role ‘R’, the following state also obtains.

\(<<n, P>, \text{playing role } R>\)

This is a state of a state.5 The following state also obtains.

\(<\text{Jones’s body, having an internal part in a state that plays role } R>\>

---

4 For presentations of this argument, see [Armstrong 1970], [Lewis 1972], [Lewis 1994] and [Braddon-Mitchell and Jackson 1996].

5 It is a hard question just how \(<<n, P>, \text{playing role } R>\) is related to \(<n, P>\). There are two reasons to think that the obtaining of the former may not supervene on that of the latter. First, on some views playing functional role R involves having a certain kind of cause. If so, and if \(<n, P>\) was caused to obtain in some other way (say by an electric probe), \(<<n, P>, \text{playing role } R>\) might not obtain. Second, on some views playing functional role R involves actually playing a certain role, as opposed to merely being capable of playing such a role. If so, then if \(<n, P>\) obtained in an environment in which it did not play that role (perhaps because of damage to parts of the brain with which it would normally interact), then it would not play role R.
This is sometimes called the *role state*.\(^6\) It is a second order state since one of its constituents is a second-order property: a property of having some property or state. Finally, the following state also obtains.

\(<\text{Jones's body, having }<n, P>>\>

It is the state of Jones’s body’s having that internal neural state.\(^7\)

Now, according to the Identity Thesis, Jones’ state of believing that water is wet simply is whatever internal state plays the appropriate causal role. Since, by hypothesis, it is \(<n, P>\) that plays this role, Jones’ psychological state just is \(<n, P>\).

\[(I) \quad \langle \text{Jones, believing that water is wet} \rangle = <n, P>\]

Now, its proponents admit that accepting (I) has some costs, but insist that these are far outweighed by its benefits. Let us consider the costs first.

**III. The Costs.**

There are (at least) two costs involved in accepting (I). Both can be seen in connection with the Principle of State Identity mentioned in section I.

A cost that proponents of this view admit to is that their view entails that beings without brains or central nervous systems cannot share our psychological states, no matter how much their behaviour might resemble ours. Recall that according to the Principle of State Identity, state A is identical with state B only if the property instantiated in A is that instantiated in B. Given this, (I) entails that the property of believing that water is wet is the property \(P\), a neurological property. I have not said much about property \(P\), but it seems plausible to suppose that only a brain, or a central nervous system, can have it. So if \(P\) is indeed the property of believing that water is wet, then only beings with brains or central nervous systems can believe that water is wet. This is a cost since many hold, and indeed it would seem to be part of our ordinary conception of the mental, that beings very different from us could share our mental states. But Lewis and Braddon-Mitchell and Jackson insist nonetheless that it is a cost worth paying.

Before discussing their reasons for this, I want to note a different, and more serious, cost. Recall that state A is identical with state B only if the subject of state A is the subject of state B. Given this, (I) entails that Jones is a neuron, or a cluster of neurones. For (I) states that Jones’ believing that water is wet just is a certain neuron’s, or neural cluster’s, having property \(P\). But this is surely false. Jones is not a neuron or a neural cluster, or even a brain. The primary cost, as I see it, with accepting (I) is that it entails that brains or brain parts are the subjects of psychological states.

This difficulty is related to an objection that was made when the Identity Thesis was first advanced. The objection, now known as the location objection, is that psychological

---

\(6\) The obtaining of \(\langle \text{Jones’ body, having an internal part in a state that plays } R>\) supervenes on that of \(<<n, P>, \text{ playing role } R>\) if, but only if, it is essential to \(n\) that it be a part of Jones’ body.

\(7\) Its obtaining supervenes on that of \(<n, P>\) only if it is essential to \(n\) that it be part of Jones’ body.

I am indebted to referees for this journal for drawing my attention to the importance of this state. Its importance will emerge in section IV.
states cannot be identical with brain states since the former are not located where the latter are. If we suppose that a state is located where its subject is, then this objection has considerable force. For since Jones is not his brain or one of its parts, Jones is not in the very same place as his brain, or one of its parts.

A possible reply is suggested in a remark of Braddon-Mitchell and Jackson.

[From a functionalist perspective, there are two options concerning the metaphysics of psychological properties. Functionalism says that x is in pain if and only if x is in a state playing the pain role, but this thesis about truth conditions for being in pain is compatible with holding, qua metaphysical thesis, either that pain is the realizer state or that it is the role state. The mind-brain type-type identity theory takes the option of identifying psychological properties with realizer states, ..., not role states, and does so in order to preserve the causal efficacy of psychological properties. [Braddon-Mitchell and Jackson, 1996, p. 101]

The reply is that while the statement ‘Jones believes that water is wet’ certainly appears to be about a person, it is made true by a state inside Jones. At the semantic or linguistic level it looks like the statement ‘Jones believes that water is wet’ is ascribing a psychological property to a person, but in fact, at the metaphysical level, that statement is made true by a state of a neuron. How our language represents a fact is one thing while the nature of that fact is quite another.

But this reply misses the point. The point is not the narrow linguistic one that it sounds odd to say that a neuron is the subject of a psychological state. This is an odd thing to say. But it is odd, I think, because it is so obviously false. The debate between the dualist and the materialist is as much over what a person is as it is over what psychological properties are. To answer the Dualist, one must show, not only that psychological properties are instantiated in a physical world, but also that they are instantiated in things that are plausibly viewed as persons. And it is simply not plausible to say that neurones are the subjects of psychological states. To insist that this is but linguistic preference simply begs the question of how states of persons are related to states of neurones. And this is but a version of the original question of how psychological states are related to bodily ones.

Part, I suspect, of what obscured this mistake is a confusion of the subject of a state with the state itself. This confusion is suggested in many ways philosophers talk about psychological states. It is commonly said, for instance, that psychological states have propositional contents, or are related to the environment, or have causal powers, when, strictly, it is the subjects of such states that are related to propositions, to the environment and that have causal powers. It is Jones who is related to the proposition that water is wet, not the state: <Jones, believing that water is wet>. It is also common to shift without comment from saying that Jones is in a certain psychological state to saying that a certain psychological state is inside Jones. But this shift is just as mistaken as the shift from

---

8 For more on this objection, see the papers collected in [Borst 1970].
9 In his very influential paper, ‘The Nature of Mind’, David Armstrong writes that ‘perhaps what we mean by a mental state is some state of the person which, under suitable circumstances, brings about a certain range of behaviour. Perhaps mind can be defined not as behaviour, but rather as the inner cause of certain behaviour.’ [Armstrong 1970, 72]. Here he shifts from saying that a mental state is a state of a person, to saying that it is a state inside a person, and so of some inner part of the person’s body.
claiming that a ball is in a state of motion to claiming that a state of motion is inside the ball. One virtue of the formal notation used here to denote states is that it reveals this shift as illicit.

IV. Avoiding the Costs

Given these difficulties with (I), one might try to avoid these costs by adopting (II).

(II) <Jones, believing that water is wet> = <Jones’ body, having some internal part in a state that plays role R>

According to (II), believing that water is wet is a second-order property of having some internal state that plays a certain causal role.\(^\text{10}\) This is a property that brainless Martians might well have. And it is one that is instantiated in something, a human body, that is a plausible candidate for being a person. (II) thus avoids (I)’s costs. But Braddon-Mitchell and Jackson resist this second-order version, arguing as follows.

What causes your arm muscles to contract is not the fact that you have a brain state of the type that typically causes your arm muscles to contract—it is the brain state type itself. The point here is the same as might be made by observing that the properties of a stone throwing responsible for its breaking a window will be the properties like velocity and mass, not its having properties that are so responsible. The consequence of refusing to identify psychological properties with neurological properties is that the psychological properties become causally impotent. [Braddon-Mitchell & Jackson, 1996, p. 100]

The difficulty with (II), they charge, is that second order properties, with which it identifies psychological ones, are causally impotent.

Their charge need not be that second order properties are wholly causally impotent. They need only charge that second order properties are causally impotent with respect to the effects of the relevant first order property. In particular, they need not claim that the stone’s second order property of having a mass sufficient to break the window is wholly causally impotent, but only that it was not causally relevant to the breaking of the window. This leaves open the possibility that the second order property might contribute to the causing of other events. Let us assume in what follows that second order properties are causally irrelevant with respect to the effects of the relevant first property.\(^\text{11}\) Given this, adopting (II) avoids the costs of (I) but only by incurring far more serious costs.

But, one suggestion might be, it is not Jones’ having some part that plays role R that is his believing that water is wet, it is his having <\text{n, P}>.\(^\text{12}\) For, one might think, it is because he has that brain state that he acts as he does with respect, say, to the perceived glass of water. But this suggestion is ambiguous. It might be the following.

\(^{10}\) For a development of this, see [Kim 1998].

\(^{11}\) This is a large assumption, whose merits I cannot discuss here.

\(^{12}\) I am indebted to referees for this journal for pointing this out.
<Jones, believing that water is wet> = <Jones’s body, having <n,P>>

But this cannot be right. For <n,P> is a state token, and this identity entails that to believe that water is wet is to have that token. Clearly, this is false. Rather, the suggestion must be that to believe that water is wet is to have a state of the same type as that token.

<Jones, believing that water is wet> = <Jones’ body, having some part in <x, P>>

On this suggestion, believing that water is wet is a property of a body, and not of some internal part of a body. This avoids what I claimed was the most serious cost of (I). But, like (I), it implies that only a being with a brain could believe that water is wet, since (by hypothesis) P is a neurological property. Still, this cost might be acceptable if the property of having some part in <x, P> is causally efficacious. And it might seem to be since similar properties, such as that of having a leaking pen, seem to be causally efficacious. In this way (III) may seem to be a neat compromise between (I) and (II).

Unfortunately, the property of having some part in <x, P> is not causally efficacious in the relevant respect. To see this, suppose that Jones has a pen that is leaking onto his shirt. The following states obtain.

<pen, leaking onto Jones’ shirt>

<Jones, having something in <x, leaking onto Jones’ shirt>>

What caused the stain? Recall that according to Jackson and Braddon-Mitchell: what caused Jones’ arm to move was not fact that he had a brain state that typically causes arms to move—it was the brain state itself. By analogy: what caused the shirt to stain was not the fact that Jones had something leaking onto his shirt—it was the pen’s leaking itself. It seems to me that if what they say about the cause of the arm’s movement is right, then so is this claim about the stain’s cause. And so the same reasoning should apply in the case of the property of having some part in <x, P>: what caused Jones’ arm to move was not his having some part in <x, P>—it was that part’s being P itself. There is, in other words, no obvious reason to distinguish the property of having some part that plays role R from that of having some part in <x, P> in terms of causal efficacy. If the former is, as they claim, causally inefficacious with respect to the moving of the arm, then so is the latter. Thus, in the end, this suggestion does not avoid the most serious cost of adopting (II).

V. The Benefit

The alleged benefit of (I), the benefit that is supposed to far outweigh its costs, is that the neurological properties of the realiser states are both causally efficacious and plausibly viewed as psychological. Such a benefit might well outweigh the costs of accepting (I). But, on closer examination, there is reason to doubt whether (I) yields such a benefit.

To see this, consider what Braddon-Mitchell and Jackson say about the case of pain.

---

13 A more specific suggestion might be that to believe that water is wet is to have some part that is of the same type as n in state <x, P>. Since my discussion will apply equally to this suggestion, I will ignore it in what follows.
[I]f... it is C fibers in us, but D fibers in dolphins, that play the pain role, then identity theorists must restrict themselves to saying: 'Pain in humans = C fibers firing' and 'Pain in dolphins = D fibers firing'. [Braddon-Mitchell & Jackson, 1996, p. 99]

This proposal is representative of those advanced by proponents of this view. The proposals specify both a region of the brain and some property of that region. Such proposals seem on their face quite plausible. But, on closer inspection, it is not clear from what they say just what the proposed identity is supposed to be. The slogan ‘pain is C fibres firing’ conflates several very different proposals.14 The phrase ‘C fibres firing’, to the right of the identity sign, is most naturally taken to denote the property of being a C fibre’s firing, which is a property of a state (or event) token. So the proposal might be to identify pain with the property of being a C fibre’s firing. But it might instead be to identify pain with the property of having a C fibre firing. Or it might be to identify pain with the property of the C fibre itself, namely that of firing. These are, I think, the only candidates. However, none is both causally efficacious in the relevant sense and plausibly viewed as correlated with being in pain.

Consider, first, the property of having a C fibre firing. Notice that this property is second-order. It is not a property of a C fibre. It is a property of something that has a C fibre. And it is just as second order as the properties of having some internal state that plays role R, and of having some part in <x, P>. All are properties of having some property or state. If, as Braddon-Mitchell and Jackson contend, second-order properties are causally impotent, then having a C fibre firing is not a property that is both causally efficacious and plausibly psychological.

What about identifying pain with being a C fibre firing? The property of being a C fibre firing is a property of a state (or event): e.g., <C fibre 21, firing>, being a C fibre firing>. One might question whether this property is causally efficacious. But even if it is, it is not a property whose instantiations would play role R. Consider a ball moving with velocity V.

<Ball, velocity V>

This state has the property of being an instantiation of velocity V in a ball.

<<Ball, velocity V>, being a ball’s having velocity V>

Surely it would be the first state that would cause the window to break (thus playing the velocity V role), not the second. Likewise, it seems, it would be the firing of the C fibre, and not that firing’s being a firing of a C fibre, that would cause Jones’ arm to move. So this proposal fails to identify a property that is causally efficacious in the right way.

14 David Lewis writes that ‘If the state of having neurones hooked up in a certain way and firing in a certain pattern is the state properly apt for causing and being caused…then that neural state is pain.’ [Lewis, 1980, 218]. But this proposal suffers from the same unclarity. Is the proposal to identify pain with the property of being a neuron’s being so hooked up and firing; or with the property of having neurones so hooked up and firing; or with the property of the neurones, namely of being so hooked up and firing?
Finally, what about identifying human pain with the property of firing? No doubt this is a causally efficacious property, if any is. But is it plausible to identify it with (human) being in pain? It would be only if firing and (human) pain are necessarily co-instantiated. But this seems highly unlikely. For neurones throughout the brain fire regularly, even when the subject feels no pain. So the property of firing, or even of firing at a certain rate, though a causally efficacious property, is not plausibly viewed as the property of human pain.

Admittedly, the proposal that pain is the property of firing is a toy proposal. A serious proposal would surely identify human pain with a far more complex neural phenomena. There are two possible dimensions to such complexity: either a more complex property, or a more complex region. Let us consider each dimension.

First, might human pain be correlated with a more complex property of a C fibre? Presumably, it would have to be some property such as that of interacting with other C fibres and other regions of the brain in a certain way. It would, in other words, have to be a property of having a certain causal history and of being apt to have certain effects in certain situations. But the more complex this property becomes, the more it would resemble a second-order property of having certain first order ones. For the causal work would be done by the first order states of the C fibre, and not by its second order state of having those first order properties. And since, by hypothesis, second order properties are causally impotent such a more complex property would also be causally impotent.

Might human pain instead be correlated with a property of some more complex part of the brain than a C fibre? Presumably, it could only be correlated with some pattern of activity of this part, such as that of interacting in various ways with other brain regions. But to correlate pain with such a pattern of activity is again to correlate it with a second order property, and so with one that is, by the current view’s guiding assumptions, causally impotent. It would, on these assumptions, be the states of that region’s parts that would do the causal work and not any state of that region itself. Again, the very reasoning that led to the rejection of (II) would thus also lead to the conclusion that the region’s states are causally impotent.

Of course, it is an empirical matter whether human pain is correlated with some neurological property. But the difficulty I am trying to indicate is quite general. It is simply that, in the case of neurological properties, the requirements for being psychological seem to conflict with those (assumed to hold) for being causally efficacious. If second order properties are causally impotent, then to be causally efficacious, a property has to be relatively simple. But to be psychological, a property has to be relatively complex. Simple properties are not correlated with psychological ones, and complex ones are functional or second order. This conflict becomes even clearer when we move from discussing a relatively simple psychological property like being in pain to a complex one like believing that parliamentary democracy is preferable to absolute monarchy.

The discussion thus far has assumed what I called in section I the Principle of State Identity. In the case where \( <x, P> \) obtains and \( <y, Q> \) obtains,

\[
<x, P> = <y, Q> \text{ if and only if } x = y \text{ and } P = Q.
\]

I have said that this is a plausible principle about state identity. It is worth noting that a proponent of the Identity Thesis has nothing to gain by denying it. Such a denial might be
modelled on a certain idea about event identity according to which Sirhan’s shooting Kennedy might be the same event as Sirhan’s killing Kennedy, even though the property of shooting Kennedy is not the property of killing Kennedy. In the case of states, the idea would be that \( <x, P> \) might be \( <y, Q> \) even if neither \( x=y \) nor \( P=Q \). If so, then Jones’ being in pain might be, let us say, a certain C fibre’s firing. That is, on such a view of states, (1) might be true, even if (2) and (3) were false, where ‘C fibre 21’ names the C fibre in question.

\[
<Jones, \text{ being in pain}> = <\text{C fibre 21, firing}>.
\]

\[
Jones = \text{C fibre 21}.
\]

\[
\text{Being in pain} = \text{firing}.
\]

The difficulty with this account of state identity is that it leaves unanswered the very question the Identity Thesis was meant to answer: How are mental states related to physical ones?

To see this, consider Jones’ state of being in pain. Let us call that state ‘S’. According to this view, S is both an instantiation of the property of being in pain, and an instantiation of the property of firing. S is an instantiation of both a mental property and a physical one. But now, how is its being the one related to its being the other? That is, how is its being an instantiation of pain related to its being an instantiation of firing? It cannot be that S’s being an instantiation of pain just is its being an instantiation of firing since, by hypothesis, being in pain is not the property of firing. But perhaps S’s being an instantiation of pain supervenes on, or is determined by, its being an instantiation of firing. Though this answers the question of how the two states of S are related, it raises the equally pressing question of whether supervening states are causally relevant to the effects of the states on which they supervene.15 But the promise of the Identity Thesis was to explain the relation of mental and physical states in a way that made the former causally relevant. Denying the Principle of State Identity thus not only leaves one without a clear conception of states, it also fails to deliver on the promise of the Identity Thesis. Better, then, to stick with the Principle of State Identity.

We were told that the costs of accepting (I) are worth paying since paying them buys causally efficacious psychological properties. But we now see that this advertisement was misleading. The costs are higher than we were told since the view conflicts not only with the idea that mental states are multiply realisable, but also with the idea that such things as persons are the subjects of psychological states. These ideas are core elements of our ordinary conception of the mental. Moreover, it is not clear that paying these costs will even deliver the goods. For, on the assumption that second order properties are not appropriately causally relevant, it is not clear that there are neurological properties that are both causally efficacious and plausibly correlated with psychological properties.

15 I do not mean to suggest that there is no possible answer to this. My point is only that it is not clear that a proponent of the Identity Thesis has much to gain, though clearly she has much to lose, by denying the Principle of State Identity.
So where does this leave us? Recall that the Identity Thesis was arrived at on the basis of two claims. The first was that psychological states play causal roles. The second was that such roles are played by states in (or of) the brain. One response is to deny the first claim. An eliminativist version would then conclude that mental states are mere fictions. A different and in my view more promising response is to deny the second claim, and to hold that psychological states are causally efficient states of whole organisms. Developing this response is, however, a task for another occasion.\textsuperscript{16}

Received: April 2000  
Revised: October 2000

REFERENCES


\textsuperscript{16} I am indebted to Andrew Botterell, Andrew Hunter, the members of the University at Buffalo Cognitive Science Group, and to two referees for this journal, for helpful comments on earlier versions of this paper.