

LYNNE RUDDER BAKER

ATTITUDES AS NONENTITIES

(Received 9 March 1994)

It is accepted wisdom among contemporary philosophers that beliefs, if there are any, are constituted by states of the brain. Underlying such a conception is the assumption that in order to explain behavior causally, beliefs must be brain states. This assumption is a materialistic descendant of Cartesian thinking. Since almost nobody today takes the idea of beliefs as soul-states very seriously, it is nearly universally assumed that beliefs must be brain states.¹ Although I agree with the Cartesian materialist that beliefs are not immaterial soul-states, I think that the conception of beliefs as brain states is badly misguided. I hope to show that "beliefs are brain states or soul states" is a false dichotomy.

I am using the phrase 'beliefs as brain states' to cover several familiar theses: the token-identity thesis, according to which beliefs are identical to brain-state tokens; nonreductive materialism, according to which beliefs are constituted by brain states (as pebbles are constituted by molecules); and functionalism, according to which beliefs are functional roles occupied by brain states. To hold that beliefs are brain states in any of these senses is to hold what I shall call 'the Standard View.'

My aim is first to undercut the central motivation for the Standard View – the assumption that the causal-explanatoriness of beliefs requires them to be constituted by brain states – and then to pose an alternative to the Standard View, an alternative that allows beliefs to be causally-explanatory without being brain states *or* soul-states.

HOW BELIEFS EXPLAIN

The major motivation for the Standard View is that it purports to show how beliefs can be causally-explanatory. In this section, I want to meet

the challenge of the Standard View by arguing, first, that the casual-explanatoriness of beliefs does not depend on their being constituted by brain states, and, second, that, even assuming that beliefs are brain states, the Standard View does not show how beliefs can be causally-explanatory in any case. In the next section, I shall suggest a different way to understand beliefs.

Before turning to my argument that it is not in virtue of being brain states that beliefs are explanatory, let me make some terminological points. (i) I take the notion of an explanation to have epistemic import. This is not to say that explanations (or correct explanations) are confined to what we now have at hand, or even to what we can now envisage. But, in the sense that concerns me here, I would not count something as an explanation if it is unknowable in principle, or if it would remain undiscovered at the idealized Peircean 'end of inquiry.' (ii) An *explanandum* is what an explanation explains. *Explananda* may be thought of as facts, propositions, states of affairs, or in Kim's terms, events as things'-having-properties-at-times, or in Davidson's terms events-described-in-certain-ways. I'll use the word 'phenomenon' as a cover term for any such *explananda*.

(iii) By *belief-explanation*, I mean not only explanations that cite beliefs, but also explanations that cite other attitudes, or a person's reasons, goals or purposes. For example, I would count "She returned to the office at midnight to shred the documents before the subpoena arrived" as a belief-explanation. (I'm using 'belief' as a general cover term.) (iv) By *intentional explanation*, I mean explanations that mention an intentional property, where an intentional property is one that cannot be instantiated in a world without attitudes. So, belief-explanations are intentional explanations. (v) By *nonpsychological intentional explanation*, I mean intentional explanations that do not cite any individual's psychological or mental states. Thus, most social, legal, economic, political and aesthetic explanations are nonpsychological intentional explanations – in contrast to belief-explanations, which are psychological intentional explanations.

(vi) By *physical explanation*, I mean explanations that cite properties taxonomic in physics. Ultimately, physical explanations would cite fundamental properties of microphysics. For whatever would impel

one to take beliefs to be constituted by brain states would impel one to take brain states to be constituted by molecular states and so on 'down' until one reaches entities that have only such properties as expressed by predicates such as 'is a neutrino.' The idea is that things have constituents, and those constituents have further constituents and so on. The ultimate constituents are those with fundamental physical properties. *Constitution* is supposed to be a very general phenomenon: rain is constituted by drops of water, which are constituted by molecules; wars (e.g., the U.S. Civil War) are constituted by battles (e.g. the Battle of Gettysburg) which are constituted by troop movements (e.g., Pickett's Charge), which are constituted by the advancing of individual soldiers, which are constituted by bodily motions, etc. Since constitution is a transitive relation, once we get started on this road to the physical, there is no stopping at the level of macro-phenomena.

Here is the argument – a conditional proof with (CP) as assumption – that it is not in virtue of being brain states (even if they are) that beliefs are causally-explanatory.²

- (CP) Beliefs have brain states as constituents.
- (P1) Intentional explanations generally are irreplaceable by physical explanations of the constituents of the intentional phenomena.
- (P2) Belief-explanations are intentional explanations.
- (P3) Belief-explanations are causal explanations.
- (P4) The causal-explanatoriness of beliefs requires beliefs to be brain states only if belief-explanations are replaceable by brain-state explanations of the same phenomena.

From (P1), (P2) and (CP), we have

- (C1) Belief-explanations cannot be replaced by brain-state explanations.

From (P3), (P4) and (C1), we have

- (C2) Beliefs need not be brain states in order to be causally explanatory.

Discharging the assumption (CP) of the conditional proof, we have

- (C3) Even if beliefs are brain states, beliefs need not be brain states in order to be causally explanatory.

The argument is valid, but are its premises true? Since time will not permit full argument for all the premises, I shall argue for the most controversial ones: (P1) and (P4).

Begin with (P1): The relevant contrast in (P1) is between intentional explanations that ignore physical constituents of intentional phenomena and physical explanations of the constituents of those phenomena. For example, an intentional explanation of why Bush pardoned Weinberger may advert to Bush's assessment of the political situation, to his own longstanding beliefs and desires, to the influence of advisers and so on. A physical explanation with which this may be contrasted – call it a *physical counterpart* of the intentional explanation – may advert to all manner of bodily motions, to marks on paper, to audible emissions and so on. The physical-counterpart explanation gives causal information about the physical events that constituted the pardon of Weinberger.

If a physical-counterpart explanation is to replace the intentional explanation, then either the physical-counterpart explanation is a 'deeper' explanation of the pardon than the intentional explanation, or the physical-counterpart explanation simply supersedes the intentional explanation, much as Copernican astronomy supersedes Ptolemaic astronomy. First, I shall argue that a physical-counterpart explanation is not a 'deeper' explanation of the intentional phenomenon; and then I shall argue (more briefly) that the physical-counterpart explanation can not simply supersede the intentional explanation. Therefore, I shall conclude, intentional explanations are irreplaceable by physical-counterpart explanations, and (P1) is true.

Consider a particular S&L that failed. An explanation of its bankruptcy in terms of its investments is an intentional explanation. Nothing is a bankruptcy or an investment in a world without complex economic practices, practices that could not exist in a world without attitudes.

So, to say that 'being an investment' is an intentional property in the relevant sense is not just to say that nothing is an investment unless it is done for a reason. (Being done for a reason is much more psychological than what is intended here.) A particular electronic transfer may be an investment even if it was made accidentally; what makes it an investment is not that it was done for a reason, but that it occurred in the context of certain financial practices. So, suppose that a single disastrous investment in condominiums and office buildings causally explains the S&L's bankruptcy. This explanation is a nonpsychological intentional explanation.

Consider a physical counterpart of such an explanation. The physical counterpart would be an explanation of the microphysical state that constituted the collapse of the S&L in terms of the microphysical state that constituted the making of the investment.³ Suppose that the fatal investment had been made by electronic transfer from point A to point B. Call the microphysical state that constituted the making of the investment a 'U-state,' where U-states are defined by the physical properties of a certain kind of stream of electrons. Also suppose that the moment of insolvency is identified with another electronic event which requested transfer of funds that the S&L did not have. Call the microphysical state that constituted this electronic request as 'F-state,' when F-states are also defined by the physical properties of a kind of stream of electrons. Now suppose that, by an extremely complex physical process, U caused F. (If this seems unlikely to you, then you are already skeptical that a physical-counterpart explanation is a deeper explanation of the bankruptcy than is the intentional explanation.)

For the purposes of this argument, I am stipulating the following: (i) an explanation of this particular S&L's bankruptcy in terms of its investment is an acceptable explanation of the bankruptcy (as acceptable as intentional explanations get); and (ii) the investment is constituted by a complex physical state that I am calling a 'U-state,' and that the event of bankruptcy is constituted by a complex physical state that I am calling an 'F-state'; and (iii) there are physical laws in virtue of which U-states cause F-states. According to the view that I want to refute, an explanation in terms of U-states is a deeper explanation of the failure of the S&L than is the explanation in terms of investments. On

the contrary, I shall try to show that, even with the given stipulations, the microphysical states are explanatorily irrelevant to the failure of the S&L.⁴

There are both epistemological and metaphysical points to be made here. First, an epistemological point: The justification for claiming that the investments caused the failure of the S&L needs no recourse to anything specifiable in terms of microphysical states. If justification of the intentional explanation depended on knowing anything about such underlying physical states, we would never be justified in claiming that the bad investments, or any other investments, ever caused anything. Indeed, we do not even know any *macrophysical* explanations without intentional presuppositions for bankruptcies. (Explanations in terms of contracts, cash-flow or even of data entries all have intentional presuppositions.)

Furthermore, although a main purpose of looking for cases of the failure of an S&L is to prevent such costly debacles in the future, knowing the underlying physical states U and F would not help prevent future S&L failures. Since any single investment may be constituted by indefinitely many different combinations of microphysical states – quadrillions, probably – a physical explanation in terms of the U-state's causing the F-state would not be “projectable” to future S&L failures: two similar S&L failures would almost surely have wholly dissimilar causal explanations in terms of microphysical states. Hence, to count the physical explanation an explanation of the S&L's collapse would simply defeat the point of explanation.

The reason that a physical-counterpart explanation would give no predictive grip on future failures of S&L's leads to the metaphysical point. Under what conditions would a physical-counterpart explanation of an intentional phenomenon be a deeper explanation than the intentional explanation? The answer concerns supervenience. The root idea of supervenience is that some properties (the supervening properties) depend on underlying properties (the subvening, or base, properties). Two things that are just alike in their base properties are just alike in the properties that supervene on the base properties. For example, if being a good person supervenes on being a kind, honest, and generous person, then two people who are just alike in terms of kindness, honesty and

generosity are just alike in terms of goodness. One of them good if and only if the other is good. So: A property P supervenes on a property P* only if, necessarily, for any instantiation of P* there is an instantiation of P⁵. Say that an explanation E supervenes on an explanation E* if and only if for any property mentioned in E, there is a property mentioned in E* on which the E-property supervenes. Now: Explanation E* is a deeper explanation of the phenomenon explained by E if and only if E supervenes on E*.

Is the explanation in terms of U-states a deeper explanation of the S&L's collapse than the explanation in terms of investments? The answer is yes if and only if the worlds in which both the U-state and the F-state obtain are worlds in which the S&L made certain investments and went bankrupt. But the supervenience relations do not hold.

To see that the relevant supervenience relations do *not* hold, notice that U-states and F-states could occur spontaneously in outer space, or in a world just like ours after nuclear war. (A few years ago, our politicians were warning us that a world like ours that had been obliterated by nuclear war was a very nearby world.) But in the absence of an economic system, such states would constitute neither the making of bad investments nor financial failure. A U-state would constitute a macrostate which in our world is an investment, and an F-state would constitute a macrostate which in our world is a bankruptcy; but *in the world without an economic system*, the macrostates constituted by the U-state and the F-state would not have the properties of being an investment and being a bankruptcy, respectively. Hence, being an investment does not supervene on the property of being a U-state, and being a bankruptcy does not supervene on the property of being an F-state. Since the property of being an investment does not supervene on the property of being a U-state, and the property of the S&L's collapse does not supervene on the property of being an F-state, the explanation of the F-state in terms of the U-state is not a deeper explanation of the S&L's collapse. To summarize:

- a. Explanation E* is a deeper explanation of the phenomenon explained by E if and only if E supervenes on E*.
- b. The explanation of the S&L's bankruptcy in terms of invest-

ments does not supervene on the explanation in terms of U-states.

Therefore,

- c. The explanation in terms of U-states is not a deeper explanation of the S&L's bankruptcy than is the explanation in terms of investments.

Intuitively, this conclusion seems to me unsurprising. Explanations always support counterfactuals. The antecedents and consequents of counterfactuals supported by the purely physical explanation do not 'track' the antecedents and consequents of counterfactuals supported by the intentional explanation. The counterfactuals supported by an explanation of the S&L's failure should include this one: If the S&L had not made the investments, it would not have failed. However, the physical explanation would only support a counterfactual like this: 'If the U-state had not occurred, the F-state would not have occurred.' But to say that the U-state did not occur (in some other world) says nothing about whether the S&L failed in that world. For, as we have seen, the occurrence of these microphysical states is neither necessary nor sufficient for the S&L's failure: The occurrence of the U-state is not necessary for the S&L's failure since the failure could have been constituted by quadrillions of other microphysical states, had they occurred instead of the F-state;⁶ the occurrence of the U-state is not sufficient for the S&L's failure since in worlds without the embedding economic systems, the U-state (and F-state) obtain without being S&L failures.

Let me venture a diagnosis of the situation. Philosophers, I suspect, have conflated supervenience (a relation among properties) with constitution (a relation among things). Let me explain. A statue is constituted by (or, some may hold, identical to) some bunch of molecules, but its property of being a statue does not supervene on the properties of the molecules. The thing is a statue in virtue of its place in the art world, or the intention of its designer, or its aesthetic qualities (such as expressiveness), or maybe something else. But none of these potential art-making properties supervenes on the properties of the molecules that constitute the statue. The properties of constituent molecules are not what makes

something a statue. The fact that the statue is constituted by molecules does not imply that its property of being a statue supervenes on the properties of its constituent molecules.

What is true of the property of being a statue is true of (wide) intentional properties generally: *supervenience* relations (among properties) diverge from *constitution* relations (among things). This point has gone unnoticed, I think, because of excessive concern for nonrelational properties of the physically described world.⁷ For nonintentional macrophysical properties, supervenience and constitution go together: A sample of water is constituted by a group of H₂O molecules, and the (macrophysical) properties of the water sample, like wetness, supervene on the properties of the molecules. But in the case of intentional properties, supervenience and constitution come apart: Even though a counterfeit bill is constituted by a certain group of molecules, the bill's property of being counterfeit does not supervene on any properties of those molecules: of two molecularly identical pieces of paper, one may be counterfeit and the other genuine. What makes a bill a counterfeit is not its physical structure. A physical duplicate of a genuine bill may be counterfeit if printed by an unauthorized person.

Apply this point to the case of the S&L. Assuming that the U-state is the microphysical state constituting the bad investment, and that the F-state is the microphysical state constituting the S&L's failure, what would make a physical explanation of the F-state in terms of the U-state an explanation of the S&L's failure? One may be tempted to say: the S&L's investments supervene on the U-state, and the S&L's failure supervenes on the F-state. But this answer harbors the fatal equivocation: the S&L's investments *are constituted by* the U-state, but the property of being an investment does *not supervene on* the property of being a U-state. This is the reason that a physical explanation of the F-state would not be an explanation of the S&L's failure, even though (by hypothesis) the S&L's failure is constituted by the F-state.

The fact that the F-state constitutes the S&L failure in the actual world depends on the social and economic institutions of the actual world. It is not in virtue of supervenience that the F-state constitutes the S&L failure; for there is no such supervenience relation between the property of being F and the property of being an S&L failure. This result

is not at odds with a suitably relaxed materialism. Although the property of being an S&L failure does not supervene on the property of being an F-state, it may well supervene on widely-distributed physical properties; if so, the relevant physical properties would be those that determine a certain kind of economic system – a whole complex set of practices – not just the local properties that determine F-states. A supervenience base for the property of being a bank failure or for the property of being an investment would include properties instantiated over vast reaches of space and time, extended far enough to include all the physical property instantiations on which economic practices depend – perhaps back to Adam Smith's day, or even to the Big Bang.⁸ Therefore, the fact that the S&L collapse does not supervene on the property of being an F-state leaves open the possibility that it does supervene on properties of some (widely-distributed) physical states.

Even if a physical explanation of the microstate that constituted the S&L's failure is not a deeper explanation of that failure, one may be thinking along eliminativist lines, like this: "Nevertheless, the physical explanation could totally supersede the intentional explanation. The relation between the physical and intentional explanations may not be one of deep to shallow, but rather one of true or false. The explanation of combustion in terms of oxygen is not deeper than an explanation in terms of phlogiston, but it supersedes the phlogiston explanation nonetheless. Why shouldn't the relation between intentional and physical explanations be like the relation of explanations in terms of phlogiston to those in terms of oxygen?"

My reply: The situations are disanalogous. First, there are reliable (though hedged) generalizations about investments that are part of a network of economic generalizations, but not about phlogiston that are part of a network of chemical generalizations. Second, phlogiston-explanations and oxygen-explanations were competitors within a single science; they could not both be correct. Investment-explanations and U-state explanations are not even obviously competitors. They find their places in very different sciences (economics and fundamental physics), and there is no *prima facie* reason that they cannot both be correct in their respective domains.

The third reason that the investment-vs.-U-state explanations are not analogous to the phlogiston-vs.-oxygen explanations is that the phlogiston- and oxygen-explanations share a single explanandum: e.g., the calx's getting heavier. But there is a sense in which the U-state explanation and the investment-explanation explain different things: The U-state-explanation explains a physical phenomenon (the occurrence of the F-state) that also happened to constitute a bank failure; the investment-explanation explains the bank failure *as* a bank failure. (To put this in the perhaps unfortunate language of essence-and-accident: The U-state explanation explains a phenomenon that is accidentally a bank failure; the investment-explanation explains a phenomenon that is essentially a bank failure.)

Fourth, there is a single explanatory interest that the phlogiston-explanation and the oxygen-explanation would have been equally able to satisfy. However, the U-state-explanation and the investment-explanation answer to quite different explanatory interests. The U-state explanation would be useless to anyone interested in bank failures *per se*; it would support no predictions about future bank failures. Since we obviously have explanatory interests in bank failures, and intentional explanations can serve such interests while physical explanations cannot, intentional explanations cannot be replaced by or made redundant by physical-counterpart explanations. I take the S&L case to be exemplary; similar arguments apply to other cases of nonpsychological intentional phenomena.

To sum up the argument for (P1) – that nonpsychological intentional explanations of intentional phenomena cannot be replaced by physical-counterpart explanations of the constituents of the phenomena: Physical-counterpart explanations can replace intentional explanations only if either the physical-counterpart explanations are deeper explanations of the intentional phenomena or they drive out the intentional explanations as true theories drive out false ones. Since intentional explanations do not supervene on physical-counterpart explanations, physical-counterpart explanations are not deeper explanations of intentional phenomena; and since physical-counterpart explanations cannot serve the predictive and explanatory purposes that are well-served by intentional explanations, physical-counterpart explanations do not drive

out intentional explanations. Therefore, physical-counterpart explanations cannot replace intentional explanations, and (P1) is established.

Now consider (P4). (P4) follows from a very general principle. Let E be an explanation that cites some property P to explain some phenomenon F.

Then, the causal-explanatoriness of P depends on the relation of P to some lower-level property P* *only if* there is an explanation E* that cites P* and E* is a deeper explanation of F than is E. Since E is replaceable by E* just in case either E is a deeper explanation of the same phenomenon that E*, or E supersedes E*, (P4) follows as an instance of the general principle. However, I really have no argument for the general principle other than intuition. So, let me try to defend (P4) in another way.

Some may object that (P4) is too strong. They may urge that the causal-explanatoriness of beliefs may require beliefs to be brain states *even if* belief-explanations are not replaceable by brain-state explanations. Instead, the objector may claim that (P4) should be replaced with a weaker (P4'):

- (P4') The causal-explanatoriness of beliefs requires beliefs to be brain states if no human being would have beliefs without having certain kinds of brain states.

I shall try to show that (P4') is false. Assuming that no human being would have beliefs without having brain states, then I can agree that having certain kinds of brain states is a *necessary condition* for having beliefs (and, of course, having beliefs at all is a necessary condition for having beliefs that are causally-explanatory). But even if having certain kinds of brain states is a necessary condition for having beliefs, it does not follow that the causal-explanatoriness of beliefs requires them to be brain states. Any time that a belief explains behavior, then all the necessary conditions for the having of the belief *ipso facto* are satisfied – whether the necessary conditions are that the believer have certain kinds of brain states, or that the believer be in a world with H₂O, or that the believer have been born more than five seconds ago. But we cannot conclude that the belief is causally-explanatory by dint of the fact that it is constituted by brain states (if it is) – any more than that

it is causally-explanatory by dint of the fact that the believer was born more than five seconds ago. Indeed, in general, necessary conditions for the instantiation of an explanatory property are often irrelevant to the explanatory power of the property. For example, from the facts that the high winds causally explain the damage to the shed, and that a necessary condition for the high winds was that the earth have an atmosphere, it does not follow that the earth's having an atmosphere causally explains the damage to the shed: the earth had an atmosphere when the shed was intact.

So, again, from the facts that beliefs causally explain intentional behavior and that having certain brain states is necessary for having beliefs, it does not follow that the causal-explanatoriness of those beliefs requires them to be constituted by brain states (even if they are). Suppose that the universe had turned out to have been Aristotelian. How would that have affected assumptions about the attitudes' being brain states? A lot: An Aristotelian (on some interpretations) would not construe beliefs as brain states.⁹ How would a supposed Aristotelian universe have affected our explanations of intentional phenomena? Not one whit: if the sensible and social world were the same as in our world, we would have had the same range of explanations, deployed in the same ways and with the same degree of success, that we actually have. In an Aristotelian universe, we would still have explanations in terms of beliefs, as well as explanations in terms of making bad investments. Thus, since the explanatoriness of belief is compatible with the world's being Aristotelian, but beliefs' being brain states is not compatible with the world's being Aristotelian, it follows that the explanatoriness of the attitudes does not require that attitudes be brain states.

In defense of (P4), I propose (but did not argue for) an intuitive general principle of which (P4) is an instance; then I rebutted a hypothetical attempt to replace (P4) with the weaker (P4') by showing that even if no human being would have beliefs unless she had certain kinds of brain states, the causal-explanatoriness of belief would not depend on that fact. Therefore, (P4) stands.

Pending defense of (P2) and (P3) – which, regrettably, I must leave for another occasion – the tentative result is that the argument that beliefs need not be brain states to be causally explanatory is sound. So,

even if beliefs (or some of them) are brain states, it is not in virtue of their being brain states that beliefs are explanatory. To say that the *explanatoriness* of beliefs (or investments) is independent of their being brain states (or computer states) is not to say that beliefs or investments are in some sense “immaterial.” They could not be – any more than our practices could be. Our practices are embodied: The macrophysical world provides brute constraints on what we do and what practices we have – e.g., we build sofas in the shapes that we do because our bodies bend that way that they do, and are comfortable in some positions and not in others – and we mold our practices around the relevant macrophysical facts.

Even if the argument for (C3) is unsound, however, I do not think that the Standard View has any account of how beliefs construed as brain states can be causally explanatory anyway. To press this point, I shall first give a general argument for the conclusion that construing beliefs as brain states would not make it clear how beliefs can explain intentional behavior; then, I shall illustrate the difficulty by criticizing a particularly impressive (but, I think, flawed) account of how the Standard View can causally explain behavior.

Suppose that I showed up at a lecture hall at 4:00 because I thought that a lecture that I wanted to hear would be delivered there then. Proponents of the Standard View would say that there was a brain state token that was my belief that a certain lecture would be delivered in that place at that time, and that that token interacted with another brain state token that was my desire to attend the lecture; and that this interaction produced further brain states that eventuated in the bodily motions that took me to the lecture hall. But on this view, the causal-explanatoriness of the token neural state derives from its *intrinsic* physical properties; assuming externalism (i.e., the view that beliefs are individuated in part by extrinsic properties, by reference to the individual’s environment), what makes the brain state a belief that *p* are *extrinsic* properties. So, the fact that the brain state was a belief about a certain lecture was irrelevant to its interactions with other neural states. On this token-identity account, the property of believing that *p* just drops out of the causal-explanatory picture.¹⁰ Hence, even if beliefs were token-identical to brain states, that fact would not show how believing that *p* could be causally explanatory.

One impressive attempt to reconcile an externalist version of the Standard View with the casual-explanatoriness of beliefs has been presented by Frank Jackson and Philip Pettit. Construing beliefs as brain states, they agree that the property of believing that *p* is not a causally efficacious property. Nevertheless, Jackson and Pettit urge, the property of believing that *p* may still be causally explanatory.¹¹ In their view, belief-explanations explain by quantifying over a range of efficacious properties (such as physical properties of brain states). Let's look at this account of how beliefs explain.

Jackson and Pettit try to show how beliefs (individuated by broad content) can be explanatory by showing how they may "programme" the phenomena that they explain without actually producing those phenomena. Programme explanations "tell us about the range of states that do or would produce the result without telling us which state in fact did the job" (396). An attribution of belief in a programme explanation picks out a range of internal states, "each member of which would have produced the result and one of which did in fact produce the result" (393).

If I am right, then this is *not* the way that belief-explanations explain. To see that belief-explanations could not explain by picking out a range of internal states, consider three questions: (i) What is the extent of the range? (ii) How can individual brain states in the range be identified by investigators? (iii) How do explanatory locutions every acquire such a range as their extension? I believe that none of these questions has a satisfactory answer. Suppose that the explanation of Smith's vote against the school budget is that he believed that it would raise taxes too much, and suppose that the attribution – 'Smith believed that the proposed school budget would raise taxes too much' – picks out a range of internal states as Jackson and Pettit claim.

First, what is the extent of the range of brain states? All the states that would have caused Smith to vote against the school budget are in the range. What states are those? If the vote had been by secret ballot, the brain states would have been all those brain states that would have caused his hand to move in any of the ways that would have produced on paper what would count as a "no" vote; if the vote had been by right-hand-raising, the states would have been all those brain states that

would have caused his right hand to go up at a time that would count as a “no” vote; and so on for all the other ways the vote could have been taken. When we attribute the belief that the proposed budget would raise taxes too much, are we quantifying over all the brain states that would have produced any movement that would have counted as a ‘no’ vote, no matter how the vote was taken, or are we only quantifying over those brain states that would have produced a movement that counted as a ‘no’ vote, given the manner in which the vote was actually taken? The very meaning of the explanation rides on how this question is answered, but it is different to see what answer could be given. It is not even obvious that there is a fact of the matter about how many ways the vote could have been taken.

Second, how could the range – assuming that there is a fact of the matter about what range it is – be identified by investigators? We must have some way of identifying the range other than as ‘any of the brain states that would have caused a motion that would be a “no” vote.’ Otherwise, recourse to programme explanations is vacuous. To see this, suppose that we disagree about why Smith voted against the school budget. I attribute to him the belief that it would raise taxes too much; you attribute the belief that the school board takes kickbacks. If the only way to identify the relevant ranges of brain states is ‘any of the brain states that would have caused a motion that was a “no” vote,’ there is no way to distinguish between your explanation and mine; the attributions would purport to pick out the same range – viz., the range that includes any of the brain states that would have caused a “no” vote. We have no nonintentional way to identify any particular brain state (much less a range of possible brain states) as being a belief that the school budget would raise taxes too much as opposed to a belief that the school board takes kickbacks. Nothing weaker than a type-identity theory – the prospects for which are dim – would yield a way to identify the range other than vacuously as the range any of whose states would have produced a motion that would have counted as a “no” vote.

Finally, how could the English ascription – ‘believe that the proposed school budget would raise taxes too much’ – have ever have come to refer to such a range of existing and nonexisting states? We attribute beliefs and successfully explain behavior in total ignorance of brain

states. Our ability to use belief ascriptions in the way that we do would be wholly mysterious if they explained by denoting brain states or by quantifying over brain states.

Proponents of the Standard View may reply like this: Just as people could use 'water' to pick out H_2O before anyone knew the chemical composition of water, so too we use 'belief' to pick out an internal state (or a range of internal states) before we know what neurophysiological states we are referring to. Such physicalists, I believe, err in assimilating 'belief' to physical natural kinds terms like 'water'. For there are lots of nontrivial ways to pick out what 'water' refers to: 'That stuff [demonstrating a body of water],' 'the liquid with the same chemical composition as this sample,' and so on. Then, later it is discovered that that stuff is H_2O . But, in the absence of type-identity, we can not similarly identify an entity as a belief token and then later discover that the entity was a particular brain state token. The analogy is misguided. (Indeed, if connectionism is true, it would seem hopeless to suppose that particular brain state tokens could be singled out as being particular beliefs. This result, I hasten to add, casts no doubt on belief, but on the Standard View that takes beliefs to be brain states.)

So, if we need some account of how beliefs explain, Jackson and Pettit's view does not fill the bill. Indeed, I do not think that any account that bases the explanatoriness of belief on the putative connections between beliefs and brain states will succeed. As I have argued, the explanatoriness of belief – like the explanatoriness of the S&L's investments – depends in part on our practices and institutions, not wholly on physical states that constitute beliefs (if beliefs are constituted by anything at all). So, if we need a philosophical account of how beliefs explain, the thesis that beliefs are brain states will not supply it. Therefore, consideration of the causal explanatoriness of beliefs provides no grounds for the Standard View of beliefs as brain states.

WHAT ARE BELIEFS?

In order to avoid begging the question against the Standard View, so far I have assumed that even if beliefs are constituted by brain states, their

explanatory power does not depend on their being so constituted. The conclusion so far is that considerations of the explanatoriness of belief provide no support for the Standard View of beliefs as brain states. Now I want to pose an alternative to the Standard View. On the alternative, a belief is a global state of a whole person, not of any proper part of the person (which as the brain).¹² Persons have beliefs; brains have states. Having certain brain states is (presumably) necessary for persons to have beliefs; but it does not follow that for a person to have a particular belief, there is a brain state that constitutes that belief. Compare: Horses win races; legs have states. Having certain leg states is (presumably) necessary for horses to win races; but it does not follow that for a horse to win a particular race, there is a leg state that constitutes the winning of the race.

The term 'belief' is a nominalization of 'believes that.'¹³ I doubt that there is a property of *being a belief that* Clinton is President, but I do not doubt that there is a property of *having a belief that* Clinton is President (or, equivalently, a property of *believing that* Clinton is President). It is importantly false that S has a belief if and only if S has a particular brain state that satisfies the open sentence 'x is a belief that p.'¹⁴

An attitude is a state in the attenuated sense in which a state of health or a state of emergency or a state of physical fitness is a state. Consider physical fitness. First, there are genuine facts about physical even though 'state of physical fitness' does not refer to any subpersonal entity. Though internal states are surely relevant to one's state of physical fitness, one's physical fitness cannot be identified with any particular internal state, nor is being physically fit a property of any internal state; one's state of physical fitness is a state of the whole person.¹⁵ Although there are borderline cases of physical fitness, there are many clear cases in which one person is physically fit, and another is not. Finally, a state of physical fitness (or the property of being physically fit) is causally explanatory in the sense that I have been exploring.

Whether a person has a particular belief (individuated by a 'that'-clause in its attribution) depends on what she would do, say, and think in various circumstances. That is, whether 'S believes that p' is true is determined by a range of counterfactuals concerning what S would do in various circumstances, where "what S would do" may itself be

specified intentionally.¹⁶ The antecedent of a relevant counterfactual may mention other of S's attitudes, but not, of course, the belief in question. If S is a speaker of a language, then the relevant counterfactuals concern her linguistic as well as her nonlinguistic behavior. These counterfactuals bear the weight of revealing the "nature" of having beliefs and the other attitudes – regardless of what is going on in the brain. (Similarly, a person is in a certain state of physical fitness if and only if certain counterfactuals are true: e.g., the person could run 100-yards in n seconds.)

Call a set of relevant counterfactuals whose truth would guarantee the truth of the attribution to a person S of a belief that p at t, a 'set of associated counterfactuals for S's believing that p at t' – for short, 'an associated set.' An associated set is a minimal set of counterfactuals true of S that guarantees that S believes that p. Not only may there be different associated sets for different people who share a belief or for a single person at different times, but also there may be more than one associated set for a single person's having a given belief at a particular time. The counterfactuals in associated sets (partially) specify conditions under which a particular person would do, say, or think certain things. To take a trivial example, "If y had borrowed x's folding umbrella and x was about to go to Seattle, where she wanted to keep dry, x would ask y to return the umbrella" may be part of an associated set for x's believing that it might rain in Seattle.¹⁷

This trivial example illustrates several features of this approach to belief. First, it is stubbornly nonreductive; it makes no attempt to give nonsemantic and nonintentional conditions for believing that p: If X is an associated set for S's belief that p at t, then the counterfactuals in X may mention S's attitudes (other than belief that p) and S's linguistic behavior understood semantically (as interpreted sentences in a public language) and/or S's linguistic behavior understood pragmatically (as speech acts). Many assume that anything less than a reductive account of the attitudes is "circular". But why? No one supposes that absence of an account of, say, goods and services in noneconomic (and to mention nonintentional) terms renders economics "circular."

The relation between the meaning of the 'that'-clause of a true attribution of belief and an associated set is rather like the relation between

the meaning 'is in excellent financial health' (said of a prudent and fortunate S&L) and a set of statements about its debt, assets, investments, cash-flow, management, and so on. (Note that this is a different kind of relation from the "constitution" relations discussed earlier.)¹⁸ Even if the statements about assets are not themselves counterfactual, their significance for the financial health of the S&L depends on counterfactuals that they support. E.g., the relevance of some asset for the financial health of the institution may lie in its liquidity, and the notion of liquidity is overtly counterfactual: An asset is liquid if you can get rid of it in a hurry at full value.

Now consider the parallels with belief: We know what kinds of statements to look for in an associated set in part by knowing the meaning of the 'that'-clause of the attribution, in part by knowing generalizations about how people behave, and in part by knowing the other attitudes of the believer. Just as there are many different ways in which an S&L can be financially healthy, so too there are different associated sets that are sufficient for believing that *p*.

Although associated sets provide insight into the nature of belief, I am not claiming to give an analysis of 'believes that *p*.' The fact that a person believes that *p* is not the same fact as the fact that certain counterfactuals about the person are true. This is so, because 'x believes that *p*' may be true of *S* and *S'* even if associated sets for *S* and *S'* have few if any members in common. Suppose that in 1992 Dan Quayle and Al Gore had very different attitudes regarding the environment, but suppose that they both believe that the Competitiveness Council inhibits environmental regulation. 'If x had the opportunity, x would disband the Competitiveness Council' may have been part of an associated set for Gore's belief, not for Quayle's. 'If x had the opportunity, x would increase the powers of the Competitiveness Council' may have been part of an associated set for Quayle's belief, but not for Gore's. Yet, they still *shared* the belief that the Competitiveness Council inhibits environmental regulation. (On my view, attribution of this shared belief does not require postulating "a similarly structured object in each head.")¹⁹

In short, *S*'s believing that *p* is an irreducible fact about *S*, and is no more mysterious than *S*'s owing \$100 in income tax, or an S&L's being in good financial health. Just as two people could not differ in

their belief unless there were other differences (viz., in counterfactuals true of each) between them, so too two S&L's could not differ in their state of financial health unless there were other differences (viz., in counterfactuals true of each) between them. But it does not follow in either case that there is any reduction of the highly complex state of affairs, easily designated by, e.g., 'The S&L's being in good financial health' or 'S's believing that p.'

Although much more needs to be said about associated counterfactuals, there are some fundamental objections that need to be turned aside. Before considering the objections, however, let me tout some of the virtues of this approach: It accords with common practice – practice that extends to every part of human life. It makes it obvious why we do not have to know about a person's brain states to know about her beliefs. By not requiring beliefs to be brain states, it explicitly abandons mentalism. By not giving conditions for believing that p solely in terms of behavior, it implicitly abandons behaviorism. (I take it to be a signal virtue that this approach shows that materialism and behaviorism are not exhaustive alternatives.)

To construe beliefs as states of whole persons is to deny that beliefs are states of brains in this sense: Suppose that, by ordinary standards of evidence, a person appears to believe that grass is green; when asked what color grass is, he replies that it is green; when asked to describe Kelly green, he says that it is the color of grass; when queried about chlorophyll, he says that it makes grass green, and so on. In this case, the person cannot be shown to lack the belief that grass is green simply by citing neurological evidence.²⁰ Even if it were the case that for every belief (token), there was a brain state (token), that fact is inessential to a belief's being a belief. What makes it the case that a person has a belief is that there be a set of counterfactuals true of the person. No account in terms of brain states could capture the class of beliefs except *per accidens*.²¹

Now let's turn to two important objections: (a) that the approach is anti-realistic, and (b) that the approach is at odds with scientific psychology.

(a) There are two places where the charge of anti-realism may be lodged. First, instead of giving a metaphysical account of belief as an

entity, I settle for truth conditions of belief attributions. Second, the truth conditions of belief attributions depend on counterfactuals, which have an ineliminable pragmatic aspect.

The account that I have given of belief does not require beliefs to be spatiotemporal entities (at least not in any ordinary sense that requires entities to have constituents); yet I maintain that the account is not anti-realistic. Ordinary entities have particular spatiotemporal locations; entities are things to which one can literally point.²² Many things that exist are not entities in this sense: the Fifth Amendment, the Office of Attorney General (even when it is not filled), Charles Keating's guilt, marriage vows, stock options, hiring practices, bigotry, atonal music. About all of these things, I am a realist.²³

The terms 'realism' and 'anti-realism' are misleading, for they have been appropriated by partisans who equate being real with being a physical entity. To put it crudely, Fs are real, on their view, if and only if there is a particular bunch of physical particles that constitute an entity that has the property of being F. But I am not prepared to cede the term 'realism' to such a partisan view, any more than I am prepared to cede the term 'metaphysics' to Hegelians (or worse, the term 'immorality' to the Moral Majority).

Well, then, what does it mean to be a realist about belief if not that beliefs are brain states? Although I cannot give an analysis of what it means to be real, here are some marks of the real: What is causally explanatory is real, and what is not observer-relative is real. I have already argued that beliefs may be causally explanatory in the same sense that investments may be causally explanatory – and as anybody with a retirement account knows, this is a robust sense of 'causally explanatory.' A sufficient condition for being real in a nonpartisan sense is to have an effect, to make a difference.

On my construal of belief, having a belief is not observer-relative. There is a difference between whether there is an associated set of true counterfactuals and whether any particular human attributor can know that there is such a set. Let C* be the set of all the relevant counterfactuals true of S that do not mention any attitude that either entails or precludes S's believing that p (e.g., regretting that p). Then, an Ideal Observer who knew C*, but who had no access to all to S's

internal states, would know all of S's beliefs. (Even if S has severe brain damage, the Ideal Observer who had no access to S's internal states, but who had access to all the relevant counterfactuals would still be in a position to know all of S's beliefs.) This appeal to an Ideal Observer is merely an idealization of the way that we know of our own and each other's beliefs – not by knowing anything about internal states but by knowing the relevant counterfactuals. Ordinary human attributors of belief may be straightforwardly mistaken – either in identifying an associated set or in assessing the truth value of the counterfactuals in it. So, belief is not observer-relative.²⁴

Counterfactuals are notoriously context-dependent. But to say that counterfactuals are context-dependent is not to say that they always lack truth-value. It is context-dependent that it is now 4:00; but there is a fact of the matter about the time for all that. Counterfactuals such as 'If you had left an hour earlier, you would have missed the storm' seem to be straightforwardly true or false. Counterfactuals about a person may leave it indeterminate whether or not a person has a particular belief: even an Ideal Observer would be unable to decide. But this feature, I think, accords with the facts of believing. In such cases, there is a fact of the matter about the indeterminacy – whether we actual observers are in a position to ascertain the fact or not.

Counterfactuals are ubiquitous in cognitive practices. The assumption that counterfactuals may be ascertainably true is embedded in the conception of a 'ceteris paribus' clause, and in intuitions about causation. It is enshrined in law and custom, and is required for our practices of praise and blame. A recent Supreme Court decision in an anti-discrimination case held that even when an employer is found to have sexually discriminated against a fired employee, the firing is permissible provided that the employee would have been fired in the absence of the discrimination. To make such a decision is to assume that there is a fact of the matter about the truth of counterfactuals, and that the truth of relevant counterfactuals may be ascertained. In practice – scientific, legal, and moral – a lot rides on such assumptions. In the Cartesian tradition, it is the task of philosophy to evaluate such assumptions and to discard those that fail to pass metaphysical muster. Since I cannot defend my rejection of this aspect of Cartesianism here, I will

say simply that I see no privileged standpoint from which to discard the assumptions without which there would be no recognizably human life. In any case, we rely on counterfactuals in many contexts that have nothing to do with attitudes; many counterfactuals in associated sets and elsewhere are ascertainably true; and the admitted indeterminacy of some counterfactuals only reflects the reality of believing.

In short, there is often a fact of the matter about whether there is an associated set true of an individual; and when there is not a fact of the matter about whether there is an associated set true of an individual, there is a fact of the matter about the indeterminacy of the individual's having the belief.²⁵ So, beliefs (not construed as brain states) are both casually explanatory and non-observer-relative. I know of no better marks of what is real than these.²⁶

(b) The second fundamental objection is that psychological research relies on (and hence indirectly confirms) the conception of beliefs as brain states. In particular, the notion that believing that *p* is a matter of tokening mental representations, or sentences-in-the-brain, may seem entrenched in psychology. In light of this consideration, it may seem idle to reject the Standard View on philosophical grounds. I have a series of replies to this objection.

First, however useful it is to posit sentences-in-the-brain in psychology, it remains an empirical matter whether or not the brain is organized along lines compatible with the Standard View. (Here I think that the eliminativists are dead right.) From the mental-representationalist's perspective, neuroscience should trump, say, developmental psychology. If neurophysiologists discover states that are "isomorphic" to intuitively correct attributions of belief, then we may well conclude that beliefs are in fact brain states – just as we may conclude that spoken sentences are in fact disturbances of the air waves. But suppose that the brain is not organized in such a way that it makes sense to suppose that there are brain states that are "isomorphic" to intuitively correct attributions of belief; suppose that there are no sentences-in-the-brain of the kind required by the mental-representation theory. In that case, mental-representationalists should conclude that there are no beliefs – since, according to their theory, if there are beliefs, they are sentences-in-the-brain. But to conclude that there are no beliefs in light of disconfirmation

of mental representations would be on a par with concluding that there is no motion in light of disconfirmation of the impetus theory. Surely, the proper conclusion would be, not that there is no beliefs, but that beliefs are not sentences-in-the-brain. This conclusion is all the more obvious when we see that the causal-explanatoriness of beliefs does not depend on their being (or being constituted by) brain states.

Second, I would predict that, even if mental-representation theories were to be disconfirmed by neuroscience, there would continue to be psychological interest in the phenomena that these theories are supposed to explain – for example, a child's capacity to pretend.²⁷ Children would still pretend even if neurophysiologists found no brain states that constituted beliefs. But it is difficult to see how even to understand pretense without presupposing belief. To pretend that this empty cup contains water seems to entail believing that it does not. Now, if (i) neuroscience fails to find brain states that can be described as beliefs, and (ii) pretense is still possible, and (iii) a capacity to pretend entails a capacity to believe, then it follows that defeat of the Standard View is not defeat of belief. Psychologists may well take heart in this conclusion.

My third point is a conjecture about why psychologists seem confident that neuroscience will not disconfirm mental-representation theory. They seem to think that the data cannot be understood without the Standard View. But – and here I tread lightly – it seems to me that psychologists simply read the mental-representation theory into the *data*. For example, the editors of *Developing Theories of Mind*, say in their introduction that between learning to talk and beginning formal schooling, children “begin to recognize themselves and others as ‘things which think,’ as things which *believe, doubt, wonder, imagine, and pretend.*” This is fine, so far, as a description of the data; but then they go on, as if merely continuing to describe the data, to say that children's new recognition “marks their coming to make a systematic distinction between the world and *mental representations* of the world,” where by ‘mental representations,’ they mean sentences-in-the-brain.²⁸ Examples could be multiplied, but this suffices to suggest that the Standard View seems required by the psychological phenomena under investigation simply because of tendentious and unnecessarily theory-laden descriptions of the data. (The term ‘representation’ seems to invite a slide

from 'representation' in an ordinary sense in which beliefs represent or misrepresent states of affairs, to 'representation' in a technical sense in which mental representations are tokened in the brain.)

Finally, there are other ways to approach the data besides from the Standard View. For example, we might think of pretense as requiring a child to have mastered enough of the language to understand counterfactuals. How mastery of counterfactuals is related to neural processes is no doubt a difficult question; but there is such a thing as mastery of counterfactuals, and psychologists rely on such mastery in collecting data anyway: the *evidence* for their theories (of meta-representations and the like) comes from asking questions like this: "If John came in the room, where would he look for the candy?" So, counterfactuals are in the theoretical picture anyway. For these reasons, I think that psychology could get along fine (and remain comfortably materialistic) without commitment to the Standard View.

CONCLUSION

Beliefs play a prominent role in our ordinary causal explanations of behavior. But the Standard View – that beliefs, in order to be causally explanatory, are (or are constituted by) brain states – is ill-motivated. Consideration of the S&L case shows that physical constitution is simply irrelevant to the justificatory, explanatory and predictive uses to which we put causally-explanatory properties of nonpsychological intentional phenomena. Therefore, the assumption that beliefs must be internal states in order to be causally explanatory is unfounded. And with rejection of this assumption goes the motivation for the Standard View. Since the Standard View has an untoward consequence – that either science will vindicate beliefs-as-constituted-by-brain states, or else there are no beliefs – we may well seek an alternative to the Standard View.

I suggest an alternative that accords with the explanatory uses to which we put beliefs. The alternative is compatible with materialism (at least in the form of global supervenience) without entailing it. Since the alternative allows beliefs to be causally-explanatory as well as to be observer-independent, it should not be charged with "anti-realism."

Nor is the alternative hostile to psychology. Thus, we have good reason not to construe beliefs, pre-theoretically, as brain states. Otherwise, in view of the vulnerability of the Standard View to disconfirmation by neuroscience, we may find ourselves at the mercy of a nonintentional version of Descartes' Evil Genius.^{29,30}

NOTES

¹ I take an event to be a token state.

² This formulation was influenced by Robert Van Gulick's criticism of an earlier version.

³ The physical explanations must bottom-out in explanations in terms of fundamental physics. The relations between macrophysical explanations (e.g., lightning caused the tree to fall) and microphysical explanations (e.g., electrons, perhaps, perhaps quarks, instantiated certain laws) are as murky as any; but for some reason, physicalists do not seem to find these puzzling.

⁴ David Owens takes a similar position. He holds that "the primacy of the physics does not entail the hegemony of physical explanation." See his "Levels of Explanation," *Mind* 95 (1989): 59–79. My objective is somewhat different from his, however. My argument against explanatory reduction is intended to open up space, as it were, for an alternative to the Standard View.

⁵ As I am characterizing it, supervenience is a reflexive, nonsymmetrical and transitive relation that allows for multiple realizability of supervening properties. Jaegwon Kim, whose articles have generated the current discussion of supervenience, defines 'supervenience' as a relation between *families* of properties. Assuming, as do many proponents of the Standard View, that any Boolean combination of properties is a property, reference to families is dispensable. But even retaining the terminology of 'families,' supervenience requires that each member of the supervening family depends on some set of properties in the base family in this sense: necessarily, if the relevant properties in the base family are instantiated, then the supervening property is instantiated.

⁶ Someone may object that had some nonF-states constituted the failure of the S&L, there would have been a different failure. But we do not individuate economic (or other social) phenomena this way. If there had been a different guest list, there may have been a different dinner party; but there would not have been a different dinner party if one of the guests had had her hair cut a millimeter shorter. There would be no explanatory value in individuating macroevents so finely, even if we could do it.

⁷ The conflation is abetted by equivocal use of the term 'physically realised by' sometimes to refer to a relation among things and sometimes to refer to a relation among properties.

⁸ I would hesitate to call anything requiring reference to such an array of properties an 'explanation.'

- ⁹ For a discussion of various interpretation, see Michael V. Wedin, *Mind and Imagination in Aristotle* (New Haven: Yale University Press, 1988).
- ¹⁰ Whence 'the problem of mental causation.'
- ¹¹ Frank Jackson and Philip Pettit, "Functionalism and Broad Content," *Mind* 97 (1988): 381–400.
- ¹² I agree with Daniel C. Dennett when he says: "The subject of all the intentional attributions is the whole system (the person, the animal, or even the corporation or nation) rather than any of its parts, and individual beliefs and desires are not attributable in isolation, independently of other belief and desire attributions." "Three Kinds of Intentional Psychology," *The Intentional Stance* (Cambridge MA: MIT/Bradford, 1987): 58. As I shall argue below, I disagree that this prevents realism about belief.
- ¹³ Many have noted that 'belief' is ambiguous. Sometimes it refers to the proposition believed, and sometimes it refers to the state of believing. My interest is in the latter.
- ¹⁴ Arguments for eliminative materialism depend on not recognizing this fact. E.g., see Stephen P. Stich, *From Folk Psychology to Cognitive Science: The Case Against Belief* (Cambridge MA: MIT/Bradford, 1983).
- ¹⁵ To say that a belief is a state of a whole person is not to say that it is a state of a whole human body: Suppose that experimenters chopped away many parts of a believer's body, but left his brain intact. Then, even in the absence of a whole body, the subject may still have beliefs – provided that relevant counterfactuals remained true to him. See below.
- ¹⁶ I am proposing a kind of "holism" about belief. In *Holism* (Oxford: Blackwell, 1992), Jerry Fodor and Ernest LePore criticize what they call "meaning holism." My holism about belief does not fall under their critique.
- ¹⁷ For a fuller discussion of associated sets, see Chapter Six of my forthcoming book *Explaining Attitudes: A Practical Approach to the Mind*, to be published by Cambridge University Press.
- ¹⁸ Even though I am now denying that beliefs are constituted by brain states without denying that investments are constituted by U-states, I still hold (P3) – that belief-explanations should be assimilated to intentional explanations generally. The basis for the assimilation is that the explanatoriness of intentional properties generally does not depend on the physical properties of constituent states, *if any*.
- ¹⁹ For different purposes, Daniel C. Dennett makes this point about belief in "Three Kinds of Intentional Psychology," *The Intentional Stance*: 55.
- ²⁰ Here is a dissimilarity between beliefs and the S&L case, it is discovered that there had been no electronic transfer (there was just a malfunction on the other end), no investment would have been made. But discovery that there had been no brain state that could be identified as a particular belief would not impugn the attribution of belief, provided that there was a relevant associated set of counterfactuals.
- ²¹ This sentence is a paraphrase of Daniel Dennett's comment on Ryle's account of intelligent action in "Three Kinds of Intentional Psychology" in *The Intentional Stance*: 45.

²² Some entities, like oceans, do not have sharp boundaries; from a microphysical point of view, perhaps no observable entities have sharp boundaries. But this consideration is irrelevant here.

²³ Certainly none of these things is a "useful fiction."

²⁴ It is logically possible (though unthinkable) that nobody ever had a belief. There would be no beliefs if there were no true counterfactuals.

²⁵ I think that indeterminacy is morally important in, e.g., assessing one's culpability for what one does.

²⁶ Although recent discussions of realism simply presuppose materialism, I take the question 'What is real?' to be prior to the question 'What is material?' (whether materialism is true or not).

²⁷ See Alan M. Leslie, "Some Implications of Pretense for Mechanisms Underlying the Child's Theory of Mind" in Astington et al.: 19–46.

²⁸ Janet W. Astington, Paul L. Harris, David R. Olsen, eds., *Developing Theories of Mind* (Cambridge: Cambridge University Press, 1988): 1. Emphases in the text.

²⁹ I am especially grateful for the incisive and helpful criticisms of Robert Van Gulick, my commentator at the 1993 Oberlin Colloquium. Thanks are also due to Fred Feldman for criticizing a draft, and to Katherine Sonderegger for many helpful discussions.

³⁰ This paper is from several chapters of my forthcoming book, *Explaining Attitudes: A Practical Approach to the Mind*, to be published by Cambridge University Press.

Department of Philosophy
University of Massachusetts at Amherst
Amherst, MA 01003
USA