--from *Explaining Beliefs*: *Lynne Rudder Baker and her Critics*, Anthonie Meijers, ed. (Stanford: CSLI Publications, 2001): 17-38.

Are Beliefs Brain States?

During the past couple of decades, philosophy of mind--with its siblings, philosophy of psychology and cognitive science--has been one of the most exciting areas of philosophy. Yet, in that time, I have come to think that there is a deep flaw in the basic conception of its object of study--a deep flaw in its conception of the so-called propositional attitudes, like belief, desire, and intention. Taking belief as the fundamental propositional attitude, scientifically-minded philosophers hold that beliefs, if there are any, are brain states. I call this conception of belief 'the Standard View.'

As readers of my book, *Explaining Attitudes*, know, I have rejected the Standard View and proposed an alternative, according to which a belief may explain a bit of behavior even if there is no particular state of the brain that corresponds to having that belief. What I want to do today is to present a direct argument against the Standard View that conceives of beliefs as brain states. I shall lay out a set of simple arguments, each of which is obviously valid, in tedious detail in order to make the structure absolutely clear. This labored presentation should make it easy for those who reject my conclusion to locate the exact points of disagreement. What I hope to accomplish with these arguments--if I cannot win you over altogether--is to make explicit a line of thought that has motivated me, at least, to seek an alternative conception of belief to the Standard View. As I defend the premises, some of my methodological convictions will become apparent. Getting clear about exactly where the controversies lie, and how important they are, seems to me a worthwhile undertaking.

The Standard View

There are two forms of the Standard View--eliminative materialism, which entails that, strictly speaking, no one has ever believed anything; and noneliminative materialism, which tries to give an account of beliefs as brain states. As I am using the term, the Standard View covers an array of well-known theories (some of which may be combined in various ways): Type-identity theories, according to which types of belief are identical to types of brain states; token-identity theories, according to which particular instances (or tokens) of belief are identical to tokens of brain states; "constitution" theories, according to which beliefs are constituted by brain states, as pebbles are constituted by aggregates of molecules; functionalist theories, according to which beliefs are causal roles occupied by brain states; and eliminative-materialist theories, according to which beliefs, if there *were* any, would be brain states.

The Standard View does not require that beliefs be construed individualistically, or narrowly. What makes a particular brain state a belief that p (say, a belief that water is good to drink) may be determined partly by the believer's relations to her environment, as so-called externalists have it, or may be determined wholly by the intrinsic properties of the believer, as so-called internalists have it. Finally, the Standard View both underlies theories that postulate a language of thought and underlies theories that do not. So, the Standard View provides the background conception of belief for an extremely wide spectrum of theories.

The minimal commitment of all these theories is this:

(SV) For all persons S and propositions p, S believes that p only if there is some neural token, n, such that (i) n has the content that p, or means that p, and (ii) S tokens n.

According to (SV), the Standard View is committed to holding that for every belief, there is a particular brain state that "realizes" that belief.[1] The Standard View holds not

[1] Realization is a theoretical relation that different philosophers construe in different ways. See Ansgar Beckermann, "Introduction: Reductive and Nonreductive Physicalism" in *Emergence or*

simply that neural mechanisms underlie or subserve mental processes, but more specifically that in order to have a belief, desire or intention that p, one has a particular brain state that is identical to, or constitutes, or realizes that belief, desire or intention.

Noneliminativist Standard Viewers, who hold that many instances of 'S believes that p' are true, perform a *modus ponens* inference on (SV). Eliminativist Standard Viewers, who hold that nobody has ever believed anything or had an attitude with propositional content, perform a *modus tollens* on (SV), and conclude that no instances of 'S believes that p' are true. But both would agree that if there *were* beliefs, they would be brain states.

An Argument Against the Standard View

My argument against the Standard View is simple; the defense of its premises, however, is more complex. Here is the master argument:

1. If any form of the Standard View is true, then either some noneliminativist theory according to which beliefs are brain states is true, or eliminative materialism is true.
2. No noneliminativist theory according to which beliefs are brain states is true.
3. Eliminative materialism is not true.

∴4. No form of the Standard View is true.

Before defending the premises, let me say informally how my reasoning goes. My argument for 2 -- that no noneliminativist theory according to which beliefs are brain states is true -- is ultimately based on an empirical conjecture about the future of neuroscience: If a noneliminativist Standard-View theory is correct, then it is an empirical theory that should be confirmed by neuroscience. But neuroscientists, I predict, will not find the relevant brain states that would confirm any noneliminativist

---

Reduction? Essays on the Prospects of Nonreductive Physicalism, Ansgar Beckermann, Hans Flohr, Jaegwon Kim, eds., (Berlin: Walter de Gruyter, 1992): 18.

Standard-View theory. My argument for 3 -- that eliminative materialism is not true -- is on a different plane: If nobody ever believed anything, then many commonplace phenomena (such as philosophy conferences) could not occur. But clearly they do occur. So, eliminative materialism is false.

Now I want to lay out these arguments in some detail. I am going to set out a simple argument, first, for premise 2, then defend the controversial premise of the argument for premise 2, then defend a controversial premise of that argument, and so on. I hope that this approach of nesting simple arguments will make the logical structure of my rejection of the Standard View as clear as possible.

Argument for 2:

Let T be any noneliminative theory according to which particular beliefs are particular brain states:

2.1 If any noneliminativist theory according to which beliefs are brain states is true, then T is true.

2.2 T is not true.

∴2. No noneliminativist theory according to which beliefs are brain states is true.

Since T can be any noneliminativist Standard-View theory whatever, there is some theory for which 2.1 is true. The argument that no noneliminativist Standard-View theory is true rests on 2.2 So, here is a simple argument for 2.2:

Argument for 2.2:

2.21 If T is true, then T is either necessarily true or contingently true.

2.22 T is not necessarily true.

2.23 T is not contingently true.

∴2.2 T is not true.

The first premise of the argument for 2.2 -- 2.21 -- is self-evident. The second premise--2.22--itself requires an argument.

Argument for 2.22

2.221 If T is necessarily true, then it is necessary that human brains are organized in the way that T claims.

2.222 It is not necessary that human brains are organized in the way that T claims.

∴2.22 T is not necessarily true.

Since T is a noneliminative theory of beliefs as brain states, T includes an account of how the human brain is organized. So, if it is necessary that T is true, and according to T, the brain is organized in a certain way, then it is necessary that the brain is organized in that way. So, 2.221 is true. But however the brain is organized, it is not necessary that it is organized in that way. Under different environmental pressures, presumably the human brain would have evolved in a different way (and still have been a human brain). So, even if the human brain is in fact organized in the way that T claims, it is not necessary that the brain is so organized. Hence, 2.222 is true.

Since 2.22 follows from 2.221 and 2.222, I take it that 2.22 is established: T is not necessarily true. This establishes the second premise in the argument for 2.2, the conclusion that T is not true. So, any particular theory that is a noneliminative form of the Standard View should be understood as contingent. That is, T is an empirical theory--as I think most Standard Viewers would agree. Indeed, at least part of the motivation for the SV is to bring belief--and every aspect of mental life--within the purview of empirical science. And many versions of the Standard View explicitly aim to be scientific theories. Now turn to the third premise in the argument for 2.2--2.23--according to which T is not contingently true either. This is where my empirical conjecture will come in. Here is the argument:

Argument for 2.23:

2.231 If T is contingently true, then T will be confirmed by neuroscience.

2.232 T will not be confirmed by neuroscience.

∴2.23 T is not contingently true.

Again, the first premise--2.231--seems uncontroversial. If noneliminative versions of the Standard View are contingent (indeed, they purport to be empirical scientific theories), then they stand subject to confirmation or disconfirmation by the relevant science, which in this case is neuroscience. It is logically possible that T be a true empirical theory that is never confirmed by neuroscience; but in that case, I do not think that anyone should believe T. We expect, rightly, that empirical theories be confirmed. So, as 2.231 implies, if T is a true empirical theory about the brain, then T will be confirmed by neuroscience. 2.232, however, is more problematic. So, here is an argument that T will not be confirmed by neuroscience:

Argument for 2.232

2.2321 If T is a type-identity theory, then T will not be confirmed by neuroscience.

2.2322. If T is not a type-identity theory (e.g, if T is a token-identity theory), then T will not be confirmed by neuroscience.

∴2.232 T will not be confirmed by neuroscience.

Since any noneliminativist theory of beliefs as brain states will either be a type-identity theory or will not be a type-identity theory, the conclusion 2.232 follows from the premises. But both premises in the argument for 2.232 need defense. Start with 2.2321: If T is a type-identity theory, then T will not be confirmed by neuroscience.

According to type-identity theories, for every belief-type, there is a type of brain state N such that necessarily, S believes that p if and only if S's brain is in a state of type N. Even relativized to species, type-identity seems way too strong. For type-identity would require that there be a single brain state such that everyone who believes, e.g., that millions died in World War II, be in that state. But that seems wrong. Suppose that as an

infant, Smith, had significant brain damage; but since it occurred when he was so young, his brain compensated for the impariment, so that different neural structures took over functions that otherwise would have been lost. So, although the adult Smith shows few signs in ordinary life of his early injury, his brain is organized in a significantly different way from, say, mine. Surely, if Smith and I read the same newspaper, we could both believe that U.S.-Japanese trade relations have deteriorated--even though there is no possibility of our being in the same brain state. Sharing a belief is not a matter of having similarly structured brains. Indeed, there are many differences among the brains of adult humans, who may share beliefs. Some left-handed people have less functional lateralization than right-handed people. So, even if we restrict type-identity theories to the species *Homo sapiens*, it is not the case that there is a single type of brain state for every type belief, such that every adult human being who has that type of belief is in that type of brain state.

Finally, if a type-identity theory were correct, then the difference between a belief that failure to de-ice your sidewalk almost always constitutes negligence and a belief that failure to de-ice your sidewalk only sometimes constitutes negligence--a difference which may be crucial in a courtroom--would have to be discernible from a neurophysiological point of view. Nothing that I have ever read about neurophysiology remotely suggests that it has detection of such differences on its agenda. For all these reasons, I do not think that any type-identity theory will be confirmed by neuroscience. So, if T is a true theory, it will not be a type-identity theory, and hence a type-identity theory will not be confirmed by neuroscience. So, I think that 2.2321 is true. Turn to 2.2322: If T is not a type-identity theory (e.g., if T is a token-identity theory), then T will not be confirmed by neuroscience. Here is an argument for 2.2322:

<u>Argument for 2.2322</u>:

2.23221 If T is not a type-identity theory, then T will be confirmed by neuroscience only if: neuroscientists in the long run are able to identify particular neural tokens as tokens of the belief that p (for any belief that p)..

2.23222 It is false that: neuroscientists in the long run are able to identify particular neural tokens as tokens of the belief that p (for any belief that p). (Empirical Conjecture)

∴2.2322 If T is not a type-identity theory (e.g, if T is a token-identity theory), then T will not be confirmed by neuroscience.

2.23221 seems obviously true. Confirmation of an noneliminative version of the Standard View that is weaker than type-identity would consist of neuroscientists' identifying particular neural tokens (of *different* neural types) as tokens of a particular type of belief. The behavioral evidence would tell the neuroscientists what type of belief is in question, and the neuroscientists would look for the neural tokens that could be said to be identical to, or to constitute, tokens of that belief-type. I do not see how anything less than actual discovery of relevant brain states to regard as beliefs would confirm T. But--and the next premise 2.23222 is my empirical conjecture--neuroscientists in the long run will *not* be able to identify particular neural tokens as tokens of the belief that p (for any belief that p).

In order for particular neural tokens to be identified as constituting tokens of a belief that p, the relevant neural tokens must have in common some property recognized by neurophysiologists--even if there is not a single type of brain state shared by everyone who has a single type of belief. In order for neuroscientists to *confirm* a noneliminative version of the Standard View, the neural tokens that are supposed to constitute tokens of a particular belief-type cannot be a complete motley. They must be nonheterogeneous: The relevant neural tokens must have something in common other than the fact that they

are all said to constitute tokens of a particular type of belief. If the brain states in question were totally heterogeneous, there would be no reason to suppose that their tokens all constituted tokens of the same belief-type. The claim of token-identity (or token-constitution) would be purely *ad hoc*. My empirical conjecture is that there will not be any salient neurophysiological feature (1) that is exhibited on each occasion on which a person manifests a belief of a certain type, and (2) that would warrant calling particular neural tokens of different types each a "realization" of that belief.

Perhaps an example will make this clearer. Suppose that, a person, call him 'Fox' got himself elected to the school board; and by all ordinary standards of evidence, Fox appeared to believe that the proposed school budget would raise taxes too much. In the board meeting, he kept complaining about rising taxes, which he had pledged to fight in his campaign; he wrote letters to the editor advocating elimination of the Latin program as too costly for the taxpayer; he joined an organization dedicated to cutting taxes; he confided to his confidential diary that he wanted a school budget that did not require any higher taxes; after the vote, he said on a TV interview that he had voted against the budget because it would raise taxes. Given this behavioral evidence, it is overwhelmingly plausible to explain Fox's "no" vote on the school budget by his belief that the proposed budget would raise taxes too much.

Now suppose that we had a total brain map of Fox's brain for a year during which he ran for the school board, voted against the school budget, gave interviews on TV denouncing the school budget for raising taxes too much, and so on. And suppose that we could pinpoint on the brain map each time at which Fox did something explainable by his belief that the school budget would raise taxes too much. His body moved in remarkably different ways on each of these different occasions. From the brain map, we could see all the ceaseless electrical and chemical activity that was going on in his brain prior to each of these actions.

My empirical conjecture is that with all this neurophysiological information, there would be no neurophysiologically salient property that was instantiated on each occasion and that could plausibly be identified as Fox's belief that taxes are too high. The various neural mechanisms would not contain a nonheterogeneous set of neural tokens that plausibly could be said to constitute Fox's belief that the budget would raise taxes too much. So, my conjecture is this: if neurophysiologists had a complete neurophysiological description of all the neural processes that controlled all the different kinds of bodily motions that constituted actions explainable by a particular belief that p, they would not find for each such neural process any particular state that could plausibly be identified in each case as the belief that p. If this is so, then T will not be confirmed by neuroscience if T is weaker than a type-identity theory, and 2.2322 is true. Since I have already argued that if T is a type-identity theory, then T will not be confirmed by neuroscience, it follows that T will not be confirmed by neuroscience at all.

Now we can work our way back up to the argument for premise 2 in the master argument, the claim that no noneliminativist theory according to which beliefs are brain states is true. Most recently, I argued that if T is weaker than a type-identity theory, T will not be confirmed by neuroscience (2.232) [This was based on my empirical conjecture.]; this "empirical conjecture" argument followed an argument that if T is a type-identity theory, T will not be confirmed by neuroscience. So, we now have the conclusion 2.232 that T will not be confirmed by neuroscience. Since T will not be confirmed by neuroscience, we have the conclusion 2.23 that T is not contingently true (2.23). Adding that T is not contingently true to the earlier conclusion that T is not necessarily true (2.22), we have the conclusion that T is not true, which is 2.2 Since 2.1 is self-evident, and 2 follows from 2.1 and 2.2, we now have 2: No noneliminativist theory according to which beliefs are brain states is true.

We now need to consider premise 3 of the master argument--that eliminative materialism is not true. Before giving my argument against eliminative materialism, I want to criticize what I think is an unsound argument for eliminative materialism; then I want to present and defend a simple argument in favor of eliminative materialism. The unsound argument for eliminative materialism is this:

a. The brain is organized as a neural net.

b. If the brain is organized as a neural net, then connectionism is true.

c. If connectionism is true, then eliminative materialism is true.

∴d. Eliminative materialism is true.

I have no quarrel with a. and b. It seems likely that the brain is something like a neural net, and that some connectionist theory may well be true.[2] This seems to me to be purely an empirical question, to be settled by cognitive and neural scientists. But c., I think, is false. Elsewhere, I have criticized an influential article that contends that if connectionism is true, then so is eliminativism about the propositional attitudes.[3] But here I just want to point out that that conditional just presupposes the Standard View: For it assumes that if there were beliefs (or other propositional attitudes), then they would be brain states. Where eliminativists go wrong is to suppose that the brain is the place to look for beliefs in the first place. Let me emphasize what eliminative materialists overlook: Failure to find the relevant tokens or types of neural states with which to identify beliefs is not *ipso facto* confirmation of eliminative materialism. Failure to find the relevant neural states would confirm eliminative materialism, *only on the assumption that the Standard View is correct.* But if attributions of attitudes are not hypotheses

[2] There are lots of controversies surrounding connectionism--Is it a theory of the mind? of the brain? is it an implementation theory?--but these issues do not matter to the point at hand. Let's assume that connectionism is a true theory of something; whatever it is a true theory of, c. is false.

[3] William Ramsey, Stephen Stich, and Joseph Garon, "Connectionism, Eliminativism and the Future of Folk Psychology," in *Philosophical Perspectives 4, Action Theory and Philosophy of Mind, 1990*, James E. Tomberlin, ed. (Atascadero CA: Ridgeview Publishing Co., 1990): 499-533. I criticize this article in *Explaining Attitudes*, 75-77.

about brain states in the first place, then no amount of brain research can confirm eliminative materialism. If, as we have seen, a noneliminative version of the Standard View were correct, then neuroscience *could* confirm the *non*eliminative version by identifying particular brain states with particular beliefs. But if--as now seems likely with the success of connectionism--the brain states recognized by neuroscience cannot plausibly be identified with particular beliefs, we cannot conclude that nobody ever had a belief. On the basis of neuroscience, we are only entitled to a disjunctive conclusion: either nobody has ever believed anything or beliefs should not be regarded as brain states. So, there is no non-question-begging argument directly from neuroscience to eliminativism.

Now I think that eliminative materialism is false, but not because I think that eliminativists are wrong about the organization of the brain. On the contrary, I suspect that they are right about the organization of the brain. But it is plainly question-begging to infer from the fact that neuroscience does not find brain states that plausibly can be regarded as beliefs to the conclusion that no one ever had a belief. A non-question-begging argument for eliminativism would have to include a defense of the Standard View. The mere fact that the brain is organized in such a way that beliefs cannot be placed in one-to-one correspondence with brain states (if it is a fact) by itself lends no support to the view that no one has ever believed anything. To lend support to eliminative materialism, that fact would have to be supplemented by an argument to the effect that beliefs ought to be regarded as brain states in the first place. Now I share with the eliminativist the prediction that a completed neuroscience will not quantify over beliefs--that is my empirical conjecture--but, by rejecting the Standard View altogether, I need not conclude that no one has ever believed anything. If beliefs are not brain states in the first place, it is hardly surprising that a science of the brain is not a science of belief.

Having dispensed with this unsound argument for eliminative materialism, let me turn to the argument against eliminative materialism. Here is the argument for premise 3:

Argument for 3:

3.1 If eliminative materialism is true, then a description and explanation of all phenomena could be complete without entailing that anybody ever had believed anything (or ever had been in any state with propositional content).

3.2 A description and explanation of all phenomena could not be complete without entailing that anybody ever had believed anything.

∴3. Eliminative materialism is not true.

Although the controversial premise is 3.2, let me say a word about 3.1 3.1 simply follows from what eliminative materialism is. Eliminative materialism is not just a view about belief. (Belief is just a stand-in for any attitude.) Eliminative Materialism is about the status of propositional content and its legitimacy in psychological explanation. So, no one who rejected theories of belief (like Fodor's and Dretske's), but who accepted, say, plans and intentions as explaining behavior would be an eliminativist if the plans and intentions were individuated by propositional content. (An intention to do A differs from an intention to do B as A differs from B.) Eliminative materialism is the view that a complete description and explanation of reality would not entail that anyone has ever had any propositional attitude. If eliminative materialism is correct, then, strictly speaking, no one ever believed or desired or intended to do anything.[4]

[4] Paul M. Churchland, a noted eliminative materialist, speaks of assigning "translational" content to aliens. "We assign a specific content, p, to one of the alien's representations on the strength of whatever assurances we have that his representation plays the same abstract inferential role in his intellectual (computational) economy that the belief-that-p plays in ours. And what goes for aliens goes also for one's brothers and sisters." ("Functionalism, Qualia and Intentionality," in *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science* (Cambridge MA: MIT/Bradford, 1989): 42-43.) This just seems confused. First, if eliminative materialism is true, there is no belief-that-p in my computational economy anyway. Second, if the alien's other attitudes are very different from mine, then it would be a mistake to assign p to the alien's representation that played the same role as some representation plays in

Now for 3.2: A complete description and explanation of all phenomena would entail that some beings had believed something. A description and explanation that did not entail that some beings had states with propositional content would leave out many if not most of the phenomena that we are familiar with, and hence would not be a complete description and explanation of all phenomena. Almost all ordinary behavior--inviting people to conferences, accepting invitations, studying philosophy, developing a new theory of motion--would be left out of any description and explanation that did not entail that anyone ever believed anything. Nothing would count as an invitation, an acceptance, pursuing a course of study, developing a new theory if no one had ever had a belief or other state with propositional content. There would be no such things as legislative or judicial phenomena. Nothing would count as debate on welfare reform in the U.S. Congress, nothing would count as a life sentence in prison, nothing would count as a world war if there were no states with propositional content. Nothing would count as being wealthy or impoverished, as being happy or being miserable, without propositional content. Nothing would count as being a felon, being a philosopher, or being governor of a state. There would be no economic phenomena, no artistic phenomena, no manufacturing, no "information highway." There would be no conferences, no scientific investigation; nothing would count as an experiment or an hypothesis if there were no propositional atttiudes. The list of phenomena that would go unrecognized by a consistent eliminative materialism goes on and on.

A determined eliminative materialist could bite the bullet and say that what we think of as social and political phenomena, and all the rest that I've mentioned, are not genuine phenomena at all. The only genuine phenomena, the eliminativist may insist, are

---

my computational economy. Churchland's essay was first published in 1981. By 1989, he was speaking of "vectorial" representation, as opposed to propositional representation. (It is unclear to me that Churchland is entitled to use the term 'representation' at all. He offers no naturalistic account of what makes a given activation vector represent a particular environmental feature.)

those that can be described and explained without implying that anyone ever had any state with propositional content. This move would be blatantly ad hoc. Moreover, it would come at a significant cost. For, as I have argued elsewhere, accepting eliminative materialism would be a form of cognitive suicide.[5] If we really tried to understand the world without presupposing propositional content, nothing would make sense. Let me just give a few examples.

Far from being better explained without appeal to states with propositional content, behavior would become unintelligible in a world without attitudes. Here are some examples: (a) Our ability to predict behavior would be miraculous. Anthonie Meijers' fingers move across a keyboard; subsequently, my fingers move across a keyboard. He then predicts that I'll be in Holland in May. I in fact am in Holland in May. Amazing! How could Dr. Meijers predict such a thing? He could not have predicted my arrival unless he assumed that, e.g., I believed that I had been invited and I accepted the invitation; and usually when people accept invitations they intend to show up at the appointed time, and so on. (b) If no one had any propositional attitude, then behavior could never go wrong. In the absence of states with propositional content, no one could ever have done anything accidentally, involuntarily, or unintentionally; nor could anyone have done anything on purpose or deliberately. (c) People's explanations of their own behavior would be uniformly false. "I fired because I thought my life was in danger" would always be false, as would "I fired because I wanted her money." (Legal and criminal proceedings would make no sense.) (d) Moral judgments, which are parasitic on propositional attitudes would make no sense. If there were no difference between believing that one was doing A and believing that one was doing B, then it would be altogether inappropriate to praise or blame a person for doing A.

[5] See my *Saving Belief: A Critique of Physicalism* (Princeton: Princeton University Press, 1987).

Far from having a deeper understanding of ourselves if we eschewed propositional content, we would have no understanding of ourselves. Here are some further examples. (a) Without propositional attitudes, there would be no distinction between lying and an honest mistake. (b) Without propositional attitudes, nothing would ever have mattered to anybody. One cannot value something in the absence of beliefs, desires, and other contentful states. (Since regret requires memory, and memory requires content, if eliminative materialism were true, we would all live without regrets--no matter what we said.) (c) Without propositional attitudes, there would be no deliberation. Our mental states would have no content at all. But there is no way that Smith could be weighing the pros and cons of pursuing graduate study in philosophy, say, without having states with propositional content. (d) Without propositional attitudes, we would not be able to make sense of our own errors. If there are no states with propositional content, then not only has no one ever believed anything, but also it has never even seemed to anyone that she has believed anything. For seeming to believe is even more contentful than is belief. Indeed, the idea of understanding--understanding anything--would make no sense without propositional content. For to say what it is that one understands requires appeal to propositional content. (E.g., we understand *that* gravity is a fundamental force.)

So, to be serious about eliminative materialism is to exchange an orderly, somewhat predictable world for one that makes no sense at all. (Indeed, without propositional attitudes, the concept of making sense makes no sense.) It is not just the commonsense world of getting and spending, of being happy or miserable, of being well-off or impoverished, of being well-paid or unemployed that would be jeopardized by eliminative materialism; but also scientific inquiry itself would become unintelligible. Scientific inquiry requires proposing hypotheses, collecting data, setting up

experiments--intentional activities all. One could not theorize at all without contentful states.

Eliminative materialists speak breezily of "successor concepts" to the concepts that eliminativism would render unintelligible. These successor concepts are to be the materials for expressing whatever is true about commonsense and the practice of science. If eliminativism is correct, then the full truth of all these matters will be expressible without invoking propositional content. But no eliminativist has given any substantive indication of how the phenomena of behavior, self-understanding and scientific inquiry can be dealt with in a content-free way.[6] And I predict that it cannot be done. Eliminativism, consistently held, is the way of unintelligibility. (Of course, most eliminativists lead their lives as if eliminativism were false; otherwise, they could hardly get to work in the morning.)

Since I do not think that we can make sense of ourselves, of each other, of the world or of our scientific research about the world without beliefs and other propositional attitudes, I would resist the ad hoc move of the determined eliminative materialist who refuses to acknowledge phenomena whose occurrence entails that some people have propositional attitudes. In that case, premise 3.2 stands, and the argument against eliminative materialism is sound. Therefore,

3. Eliminative materialism is not true.

If 1 is true "by definition," and 2 and 3 have been established, we should now conclude that

[6] Paul M. Churchland has used connectionism to try to give an account of theories and explanation. Throughout, he conflates views on the nature of knowledge and views on the mechanisms that encode it. Connectionism, if true, may falsify sentences-in-the-brain models of internal mechanisms, but all that would follow is that propositions and propositional attitudes should not be understood in terms of sentences-in-the-brain. Throughout, the (plausible) claim that if connectionism is true, then sentences-in-the-brain models are false is elided with the distinct (and implausible) claim that if connectionism is true, then knowledge is nonpropositional. [This footnote is taken from my review of Churchland's *A Neurocomputational Perspective*. The review appeared in *The Philosophical Review* 101 (1992): 906-908.]

∴4. No form of the Standard View is true.

This completes the line of reasoning that leads me to deny the Standard View and look elsewhere for an understanding of belief and other attitudes. Since the argument is valid, anyone who favors the Standard View--for whatever reason--must reject at least one of the premises. By laying out the arguments as I have, I am inviting those who disagree with the conclusion to identify the premise (or premises) that they find dubious.

Two Objections to Denial of the Standard View

One may be motivated to try to find a premise to reject by the suspicion that the price of denying the Standard View is too exhorbitant. I now want to respond to two potential grounds for such suspicion: First is the charge that to deny the Standard View is to reject the relevance of neuroscience to understanding behavior. Second is the charge that to deny the Standard View is to make it impossible that beliefs causally explain behavior.

I. Some may charge that to deny the Standard View is to reject the relevance of neuroscience to understanding behavior. To that charge, I respond with an emphatic NO. My point is not that the brain is irrelevant to behavior, but rather that the relations between brain activity and propositional attitudes are much more complicated than the Standard View can allow. This, again, is my empirical conjecture. Of course, there are underlying mechanisms in the brain that "subserve" our mental processes. But--if my empirical conjecture is correct--it is not the case that for every salient element of a mental process, there is a salient element of a neural process.

Of course, I agree that psychopharmacology, even before Prozac, was making strides in controlling moods. Everybody knows that changes in the brain (after ingesting LSD, say, or two liters of beer) make for changes in behavior. But even if neurophysiology and psychopharmacology could predict when someone will start having

paranoid beliefs in general, I am doubtful that the difference between believing that one's neighbor is a space alien and believing that one is being followed by a federal agent can be detected by neurophysiology--and the difference in beliefs is important if we are trying to understand someone's behavior.

Let me try to fill this out a bit. I would expect neuroscience to tell us some things about our mental life, but not others. I would expect neuroscience to tell us about states of mind like paranoia, euphoria, dejection, confusion. But that's different from telling us why you're dejected about your tenure review. As far as we know, we cannot tinker with brain states in order to produce beliefs about tenure in someone who has never heard of tenure. I doubt that at anytime in the future, I could go in for a "brain state adjustment" and come out with the ability to speak Chinese or to write a symphony. Much of our mental life is relational and specific to culture--even though what is culture-specific takes place against a background of broader mental patterns like paranoia, euphoria, etc. I expect neuroscience to illuminate these broader mental patterns (like paranoia and depression); I am less sanguine about neuroscientific illumination of the vast portions of our mental life that are specific to culture.

Here is an analogy: Suppose that I am a fan of the Western television show, *Gunsmoke*, the longest-running television series in American TV. I now want to understand the relation between Matt Dillon, the protagonist, and the bartender, Miss Kitty. How far do I get by examining the wires and circuitry in the TV when I remove its back? Perhaps I understand why some days when I watch, Matt Dillon looks a little greenish; or perhaps I understand why some days the images flicker, and I have no sound. Perhaps I even understand the origin of the beams that give rise to what appears on the screen. But I do not understand why Miss Kitty waits for Matt Dillon for all those years. I can understand about color, sharpness and vertical hold by understanding the mechanisms inside the set; similarly, a neuroscientist can understand about mood,

alertness, and sense of balance by understanding the mechanisms inside my head. But just as we would not expect to understand Miss Kitty and Matt Dillon by understanding the TV's "insides," so too we (or at any rate, I) should not expect to understand why Smith went to law school by looking at Smith's brain.

The reason that we do not understand why Smith went to law school by looking at Smith's brain is that the attitudes that causally explain Smith's going to law school are part of a pattern at a higher level of organization than patterns exhibited by Smith's brain states. And this is where I depart from the Standard View--if I am right, there is certainly no *requirement* that there be a one-to-one correlation between the elements of intentional patterns and the elements of nonintentional patterns. And my conjecture is that in general intentional patterns *in fact* are not isomorphic to nonintentional patterns.

In some cases, it is obvious that intentional patterns of action are not mirrored by nonintentional patterns of bodily motions. Suppose that a company's auditors are looking for an embezzler. They look for patterns of moving money around from different accounts, for patterns of withdrawals and so forth. It would be astonishing if these intentional patterns were mirrored by nonintentional patterns of the embezzler's bodily motions. Similarly, intentional mental patterns in, say, deliberation, need not be mirrored by nonintentional neural patterns detectable by neurophysiologists. Having certain kinds of brain events is necessary for deliberating, but it does not follow that we should regard each propositional attitude that is a step in the deliberation as just such a brain event.

The bearer of a mental life--the deliberator, the agent--is the person, not the brain; nevertheless, a person is constituted by a body. So, it is not surprising that bodily states--like low blood sugar, poor circulation, even physical fitness--affect our mental life. And since the brain plays a crucial role in governing the body, it is not surprising that changes in the brain induced by drugs, legal and illegal, change moods and affect people's

judgment. Indeed, neurophysiology could falsify particular explanations of actions in terms of beliefs--by the discovery, say, that the subject has a brain tumor, or Alzheimer's disease. But all of this is compatible with absence of correlation between beliefs and brain states. My point, again, is not that neurophysiology is irrelevant to understanding human activity, but rather that it does not have the relevance assumed by the Standard View.

II. Second are worries about causation. How is it possible that beliefs are causally explanatory if they are not brain states? This question has the form--How is it possible that p?--of an age-old philosophical question. In the Theaetetus, Plato has Socrates ask, "How is it that we can have false belief?" Socrates admits to great perplexity in not being able to say how false belief arises in us.[7] It would have been ludicrous for Socrates, on finding himself unable to give a satisfactory account of false belief, to have concluded that there must be no such thing. Similarly, assuming that beliefs are not brain states, it would be ludicrous for us to suppose that we must give a satisfactory account of how these beliefs *could* be causally explanatory, in order to be justified in taking them to be causally explanatory.

Our only access to causal efficacy is by means of successful causal explanation; and there is no doubt that our only reliable patterns of explanations of behavior invoke attitudes. We change people's behavior by changing their attitudes. Politicians spend millions trying to produce particular attitudes in the citizenry, and the attitudes causally explain the outcomes of elections. A job seeker tries to induce favorable attitudes in his interviewer. *Mens rea*, a matter of attitude, is an essential element in the criminal law. There is simply no doubt that attitudes are causally explanatory--whether beliefs are brain states or not. Beliefs have been used in successful causal explanations for

[7] See Gareth B. Matthews, "Perplexity in Plato, Aristotle, and Tarski," *Philosophical Studies* 85 (1997): 213-228.

millenia--long before anyone conceived of them as brain states. So, our knowledge that beliefs are causally explanatory does not depend on our ability to answer the question "How is it possible that they are causally explanatory?"

Moreover, the question "How is it possible that beliefs are causally explanatory?" is not necessarily answered by assuming that beliefs are brain states. The problem of mental causation was raised almost a decade ago in articles with names like "Mind Matters" and "Making Mind Matter More" and "More on Making Mind Matter."[8] These articles assumed that beliefs *were* brain states. The problem was this: How could the fact that a brain state was a belief be relevant to what that brain state caused? Wouldn't that brain state have had the same effects--caused another brain state or caused a bodily motion--if it had not been a belief? So, even if beliefs are brain states, there would still be the question of how beliefs qua beliefs could be causally relevant to behavior? So, construal of attitudes as brain states does not necessarily solve any questions about mental causation. If you are worried about mental causation, then token-identity of beliefs and brain states is too weak to help. And type-identity, or property-identity of beliefs and brain states is, I think, totally implausible. As I have already said, it seems highly unlikely that everyone who believes that the Cold War is over instantiates the same neurophysiological property.

Worries about mental causation as they are usually expressed lead to an impasse. This suggests to me that we should reconsider the reasoning that led us to those worries. As I argued in *Explaining Attitudes* and in "Metaphysics and Mental Causation," worries about mental causation presuppose a faulty model of causation, one that does not fit actual successful explanatory practice. But not everyone agrees with my invoking pragmatic considerations in a metaphysical discussion. So, let me take another tack here

[8] See Ernest LePore and Barry Loewer, "Mind Matters," *Journal of Philosophy* 84 (1987): 630-42; Jerry Fodor, "Making Mind Matter More," *Philosophical Topics* 17 (1989): 59-80; Ernest LePore and Barry Loewer, "More on Making Mind Matter," *Philosophical Topics* 17 (1989): 175-191.

and tell a speculative story about how beliefs could be causally explanatory without being brain states.

To have a belief that p is to be ready to do, say or think various things in various circumstances. But if one is ready to do, say or think various things in various circumstances, then the brain too has a set of dispositions. Consider Jones, who wants to rise in the social world, and believes that becoming well-known in the community is the best way to improve his social status. And suppose that associated with that belief are various counterfactuals like: If x were invited to speak at a men's club luncheon, then x would accept the invitation. If x had a chance to be a conspicuous contributor to a celebrity charity, then x would give a lot of money to that charity. And so on. Now if these counterfactuals are true of Jones, then Jones's brain must have its own set of dispositions at yet a lower level. E.g., if in certain circumstances, Jones were to accept an invitation to speak at a men's club luncheon by saying "yes," then his brain would have to be disposed to move his mouth in that way in those circumstances. And if in other circumstances, Jones were to give a lot of money to a charity by writing a check, then his brain would have to be disposed to move his hand in a check-writing way in those circumstances. When Jones does these things, his brain is moving his body in certain ways. For Jones's body to move in the appropriate ways, further counterfactuals are true--this time, not of Jones, but of Jones's brain: If Jones's brain received such-and-such sensory input, it would process it in certain ways (in speech centers--Broca's area and Wernike's area), and ultimately it would make the mouth move in certain ways. And if Jones's brain received sensory input of another kind, it would process it in another way, and ultimately it would make the hand move in certain ways.

My speculation is this: From a neurological perspective, there may be no salient similarity between Jones's giving a lot of money to a charity and Jones's accepting an invitation to speak at a men's club. But from the perspective of attitudes, the episodes

are elements in a single pattern. The pattern is there. It is not just a matter of our interpretation. The episodes would not have occurred if Jones had not had the relevant beliefs (with the associated counterfactuals). But the pattern is invisible from the point of view of neurophysiology. That is my speculation. And whether the speculation is correct or not, I think that this fantasy shows how it is possible for beliefs to be causally explanatory if they are not themselves brain states.

Both the behavioral pattern and the elements in it are causally explainable by Jones's attitudes. Someone may object that since, on my view, it is logically necessary that attitudes are connected to actions by means of the associated counterfactuals, attitudes cannot causally explain actions. Of course, I agree that causes are not connected to their effects by logical necessity. However, a belief can still causally explain an action as long as the having of that belief does not depend on the performing of that action. For example, the belief that one could improve one's social status by becoming a community leader can causally explain one's accepting an invitation to speak at a men's club so long as there are counterfactuals sufficient for having that belief that make no reference to the action to be explained. For in that case, accepting the invitation would be logically independent of having the belief.

On this sketch, beliefs are causally explanatory, since if Jones had not had the belief that he could improve is social status by becoming well-known in the community, or that he could become well-known in the community by becoming a conspicuous contributor to a celebrity charity, then he would not have done the various things. His brain would not have had its dispositions to cause certain bodily motions in certain situations. But when his body moves in these various ways in various situations, it is altogether possible that entirely different brain states are engaged on different occasions. And--if there are no brain states in common to these episodes that can plausibly be

identified as Jones's beliefs--then neurophysiological explanations of Jones's various attempts to gain social status will miss the causal pattern that belief-explanations capture.

Now I admit that there are huge empirical questions about how a brain acquires its various dispositions to move a body in ways that exhibit intentional patterns of action. What I think is really amazing is that the brain's dispositions to move the body in ways appropriate to belief in various circumstances are open-ended. In new situations, the brain moves the body in ways that continue the intentional pattern. How the brain accomplishes this, I do not think that anyone knows. But this is not a philosophical question. If anybody discovers the answer, it will be scientists, not philosophers.

But the important point is the distinction between causally explanatory properties (like believing that becoming well-known in the community is the best way to improve one's social status) and whatever underlying neural mechanisms produce bodily motions that constitute actions explainable by that belief. This distinction between causally explanatory properties and underlying (physical) mechanisms is taken for granted in other areas--e.g., in economics. We say that the decline in new housing starts was caused by the rise in the discount rate (the rate that the government charges banks to borrow money). And we can tell an intentional story in economic terms about the connection between the rise in the discount rate and the decline in new housing starts, but we do not know what nonintentional, physical underlying mechanisms sustain the connection. But nobody worries that we know of no nonintentional, physical underlying mechanisms, because there is a robust causal pattern exhibited by these economic phenomena. Moreover, I would be surprised if anyone thought that, in order for the rise in the discount rate to be causally explanatory, it had to be identified with some particular state of an underlying physical mechanism salient from the point of view of physics. For the same reason, beliefs need not be identified with brain states. So, I think that the

distinction between causally explanatory properties and underlying mechanisms is useful for seeing how attitudes can be causally explanatory without being brain states.

This talk of different explanatory patterns will not sit well with those who take the task of philosophy to show how all phenomena fit into a single causal structure of microphysics-- "one size fits all." There is, I believe, a deep methodological divide between many proponents of the Standard View and the proponents of what I've called Practical Realism. So, let me conclude with some remarks about philosophical method. To make the focus as sharp as possible, I'll baldly set out what I take to be methodological maxims of both positions. My formulation of these maxims is very crude and subject to correction and refinement. I'll call the two positions 'methodological physicalism' and 'methodological pragmatism.'

The methodological physicalist starts with a theoretical picture based on a philosophical idea of fundamental physics. He looks to see what general principles--like the causal closure of the physical and strong supervenience--that picture implies. Then, for any putative kind of phenomena, he checks to see how it fits the picture. If he cannot imagine how some putative phenomena fit in the microphysical world, then it is deemed unsuited for "serious science." And according to the methodological physicalist, nothing unsuited for serious science can play an ineliminable role in a complete description and explanation of all phenomena.

The methodological pragmatist, by contrast, starts with successful explanatory practice--in everyday life as well as in the sciences--without any a priori restrictions on what is or is not suited for science. The fact that something is indispensable for successful explanatory practice (e.g., attitudes) suffices to secure it ontological status. A methodological pragmatist thinks that one's theories and one's actions should be congruent--and would think it dishonest to deny in theory what he must manifest in action. No purely metaphysical reasoning (as opposed to empirical information about

the circumstances) would make a pragmatist doubt that what she heard was caused by what the speaker said, where what the speaker said is identified by propositional content. What is right before our eyes takes precedence over metaphysical theories.[9]

If we have overwhelming reason to hold that beliefs causally explain behaivor, and good empirical reason not to identify beliefs with particular brain states, and, say, a theory of causation that requires beliefs to be brain states in order to be causally explanatory, then the methodological pragmatist says that the theory of causation should yield before successful explanatory practice. The methodological physicalist, by contrast, would hold on to the theory of causation that conforms to his metaphysical picture, and either argue that beliefs really are brain states (the noneliminativist Standard Viewer) or that beliefs do not really explain behavior (the eliminativist Standard Viewer). Both of these Standard-View strategies adjust the phenomena to his metaphysical picture; whereas the methodological pragmatist seeks to adjust his metaphysical picture to the phenomena.

It is this difference in strategy, I think, that makes the controversy over causation so difficult to settle. Both the physicalist and the pragmatist are rational, but neither is moved by the other's arguments. To the methodological physicalist, the pragmatist looks shallow and unprincipled; to the methodological pragmatist, the physicalist looks rigid and out of touch with reality. Hence, the impasse. Or at least this is the way that I see the difficulty right now. (I'd be interested to hear what others had to say about this.) For these reasons, I do not believe that I am able to refute my metaphysical opponents, but I do hope at least to have opened the door to another position.

Conclusion

[9] I would go so far as to take the fact that, e.g., the conjunction of strong supervenience and the causal closure of the physical has the consequence that all apparent macrocausation is illusory to be a *reductio ad absurdum* of the conjunction. See my "Metaphysics and Mental Causation" in Mental Causation, John Heil and Alfred Mele, eds. (Oxford: Oxford University Press, 1993): 75-95.

To sum up: I have presented a valid argument against the conception of the attitudes as brain states, and I have defended the premises--and then defended the defenses of the premises. The Standard View of the attitudes as brain states, in both its eliminative and noneliminative versions, I argued, is false. Along the way, I criticized an unsound argument for eliminative materialism. A proponent of the Standard View, I admitted, might be motivated by either of two suspicions to find one of my premises to reject: First, one may suspect that if beliefs were not brain states, then neuroscience would be irrelevant to explaining behavior; second, one may worry about how beliefs could be causally explanatory if they were not brain states. I tried to allay both worries. In any event, if one takes such worries to be reasons to endorse the Standard View, then one will have to reject at least one of the premises in my argument against the Standard View. Finally, I contrasted two approaches to philosophy, which I dubbed methodological physicalism and methodological pragmatism. Needless to say, I find myself in the pragmatist camp. The sublime elegance of physicalism is seductive, but the rough-and-tumble of pragmatism seems closer to reality as we all know it.[10]

Lynne Rudder Baker
University of Massachusetts/Amherst
Slightly Revised July 3, 1997

[10] Many thanks to Katherine A. Sonderegger for her tireless help, and to Gareth Matthews for important comments.