

--from *Explaining Beliefs: Lynne Rudder Baker and her Critics*, Anthonie Meijers, ed. (Stanford: CSLI Publications, 2001): 183-218

Practical Realism Defended: Replies to Critics

Lynne Rudder Baker

I wish to begin by thanking the contributors to this volume, who have put so much energy into reading my work, criticizing it and making helpful suggestions. Instead of engaging in tedious point-by-point combat, I shall organize my replies around a number of topics that, I trust, will be of interest to the general philosophical community. In this way, I hope not only to answer criticisms and to clear up misinterpretations, but also to further philosophical discussion about the attitudes. One dominant theme, shared by a number of my critics, is this: Belief-explanations, and macroexplanations generally, “cannot just stand on their own,” as Meyering (ms, 10) put it. I shall discuss variations on this theme under several of the topics.

The topics that I shall consider are these: (1) Causal Explanatoriness of the Attitudes (Dretske, Elugardo); (2) The “Brain-Explain” Thesis and Metaphysical Constraints on Explanation (Antony, Elugardo); (3) Causal Powers of Beliefs (Meyering); (4) Microreduction (Beckermann); (5) Non-Emergent, Non-Reductive Materialism (Antony); (6) The Master Argument Against the Standard View (Dretske, Antony, Elugardo); (7) Practical Realism Extended (Meijers); (8) Alternative to Both the Standard View and Practical Realism (Newen).

Before turning to these topics, however, let me clear up a point on which Dretske misinterprets Practical Realism and its difference from the Standard View. According to Practical Realism, believing that *p* is in the first instance a complex property of a person;

what makes it the case that a person believes that *p* is determined entirely by what that person would do, say or think in various circumstances. The term ‘belief’ is a nominalization of ‘believes that.’ Beliefs are not entities that are located at one place “inside” or “outside” a person. So it is misleading to say, as Dretske does, that according to Practical Realism, beliefs are “outside where the observable facts are that are used to tell what people believe and desire.” (Dretske, ms,1) The question of *where* a belief is arises only if one (mis)takes beliefs to be entities, as Standard Viewers do. Although Dretske’s “observable facts” are partly constitutive of S’s believing that *p*, conditions inside S’s skin may also be partly constitutive of her having that property. A person’s property of believing that *p* is a relational property “located” where the person is—just as the relational property of promising to do *A* is “located” where the promiser is. In neither case—the case of believing that *p* or of promising to do *A*—does it follow that one’s internal states are irrelevant to one’s having the property in question.

The main difference between Practical Realism and the Standard View is that, according to the Standard View, each belief has a specific location in the brain. That is, on the Standard View, for each instance of a belief, there is a particular neural configuration or token (perhaps “distributed”) that is identical to, or constitutes, that belief. Roughly, the externalist Standard Viewer takes a neural token to be a belief that *p* partly in virtue of its relations to features of the environment; whereas the internalist Standard Viewer takes a neural token to be a belief that *p* solely in virtue of the intrinsic properties of the believer. Although the Practical Realist agrees with the externalist Standard Viewer in “locat[ing] the facts that make mental states mental in [a] network of external facts,” the Practical Realist disagrees with both the externalist and the internalist Standard Viewer in supposing that each belief has a location in the brain. According to any version of the Standard View, beliefs have tokens that are spatiotemporal entities.

According to Practical Realism, there are no such *things* as beliefs; rather, believing that p is a relational property of persons.

Of course, the mere fact that believing that p is a relational property does not by itself prevent beliefs from having specific locations in brains. As Dretske points out, what makes the paper in my pocket money are its relational properties; but it is located in my pocket nonetheless. He asks: Why “can’t we locate the mind in the body (just as we keep our money in our pocket), but locate the facts that make mental states mental...in that network of external—including historical—facts that exist (partially, at least) outside a person (at the time he has the belief)?” (Dretske, ms, 5) I do not lack an answer. My answer is that, we can say that our minds are in our bodies (in some metaphorical sense of ‘in’); but our mental states do not have specific locations in our bodies—in our brains or kidneys, say. Although minds are like money in that they are what they are in virtue of their relational properties, minds are unlike money in another respect: whereas we can identify a particular piece of metal as a coin of a particular value, we have no way to identify a particular internal-physical-state token as a particular belief token. One’s mind (or brain) is not like one’s pocket, full of discrete items that can be identified in terms of relational properties like being a U.S. quarter. Now on to the actual arguments.

(1) The Causal Explanatoriness of the Attitudes. (Dretske, Elugardo)

Dretske takes me to task for my attempts to deal with how attitudes causally explain behavior. He argues that beliefs must be brain states in order to explain behavior. He says that “beliefs are states that explain the behavior of the person they are states of, and there is no way they can do this unless they exist in (or are states of) the motor control centers—the brain.” (Dretske, ms, 6) For reasons that I give in Part II of *Explaining Attitudes*, I think that Dretske’s point rests on an erroneous conception of causal explanation. There are no metaphysical constraints on causes. Causes are what

successful causal explanations cite. Although I have no analysis of ‘causal explanation,’ I do provide a bare sufficient condition for an occurrence of one property to explain the occurrence of another property. (*Explaining Attitudes*, p. 122). In general, ‘x’s having F causally explains y’s having G (in the circumstances)’ does not entail that the instantiation of F be locatable in some particular place inside x. In particular, ‘S’s believing that p causally explains S’s doing A (in the circumstances)’ does not entail that there be anything locatable inside S that *is* a belief that p. To say otherwise, I think, is covertly to place a gratuitous metaphysical constraint on what can be a cause.

Dretske continues: If beliefs are to “control” bodily movements, then “you need them *inside* with their hand on the steering wheel.” (Dretske, ms,6) But everything here depends on what is meant by ‘control.’ I take control to be subordinate to causal explanation. So, if beliefs causally explain (certain) bodily movements, then ipso facto beliefs control those movements. And S’s belief causally explains a bodily motion if the following is the case: If S had not had the belief, then S’s body would not have moved as it did; and given that S had the belief, then it was assured in the circumstances that S’s body moved as it did.

To say that a belief explains Clyde’s going to the refrigerator is not to say that there is a particular entity in Clyde’s brain, called a ‘belief,’ that reaches out into the body, so to speak, to pick up and set down legs in the direction of the refrigerator. Rather, a belief that there’s beer in the refrigerator explains Clyde’s behavior when the behavior is a manifestation of a disposition (e.g., If x wanted a beer, then, other things being equal, x would go to the refrigerator) such that (a) it is one of a set of dispositions that constitute Clyde’s believing that there is beer in the refrigerator, and (b) Clyde would not have gone to the refrigerator (other things being equal) if he had not had a set of dispositions that constitute his believing that there is beer in the refrigerator. A necessary

condition for Clyde to go to the refrigerator is that his body move in a certain way, and a necessary condition for Clyde's body to move in the relevant ways is that there be certain neural processes that produce the relevant motions. But these necessary conditions would not have obtained if Clyde had not believed that there was beer in the refrigerator.¹

Dretske thinks that the account that I give of how attitudes explain behavior at the end of "Are Beliefs Brain States?" is circular. (Dretske, ms, 10-11) So, let me try again. According to the Standard View, if a desire to improve one's social status causally explains one's writing of a check, then one was in a brain state at the time of the writing of the check that just *was* the desire to improve one's social status. I can (and do!) deny this conditional; but I need not (and do not) deny that brains produce bodily motions. If a neural process causes a bodily motion that constitutes, say, a writing of a check, there need be no distinct element in that neural process that is identifiable as a desire to improve one's social status. Yet, on my view, the causal explanation of one's writing of the check may well be that one wanted to improve one's social status.

Let me explain. If one had not wanted to improve one's social status, one's brain would not have moved one's hand in that way at that time. Now the huge unsolved problem is how a person's intentional dispositions (e.g., to improve one's social status) are coordinated with a person's brain's dispositions to go into certain sequences of states (e.g., states that move the hand). Somehow, by mechanisms that I do not think are at all understood, there is coordination of dispositions of the person and dispositions of the brain. A person's disposition to write checks in certain circumstances in order to improve social status is tied to a disposition of the person's brain to move the hand in certain ways in certain circumstances. Even though we have no account of the

¹ I spell this out in much greater detail in "What We Do: A Nonreductive Approach to Human Action" in *Human Action and Causality*, Jan Bransen and Stefaan Cuypers, eds. (Dordrecht Holland: Kluwer Academic Publishers, in press).

connection between these disparate dispositions, we do know two things: (a) we know that such coordination exists since part of the criterion for wanting to improve one's social status is that one be disposed to do what one thinks will improve one's social status in various circumstances, and one can't be disposed to do what one thinks will improve one's social status in various circumstances unless one's body moves appropriately in certain circumstances; and (b) however the coordination is effected, it does not require that there be any particular brain state that can be identified with a desire to improve one's social status.

Now let me turn more specifically to the charge of circularity. First, since I am not trying to give "a reason to introduce attitudes," there is just no space for circularity. Attitudes, in my opinion, are not "introduced," as, say, electrons are. Attitudes are part and parcel of the entire commonsense framework by means of which one makes it through the day. Second, I want to respond to the charge that "the pattern we need the attitudes to explain is only a pattern that is created by the attitude explanation." (Dretske, ms, 11) To see that this is not a consequence of my view, consider our attribution of character traits. Suppose that inspection of Frank's verbal behavior reveals a pattern of telling the truth, even when it would be to Frank's advantage to lie. We may explain this pattern by saying any of the following: Frank has a policy of not saying what he believes to be false, or Frank is just trying to win our confidence so that he can defraud us later, or Frank doesn't realize when it is to his advantage to lie. But the behavioral pattern of telling the truth is not "created by the attitude explanation." The behavioral pattern is there independently of any explanation. Indeed, there is room for rational disagreement about the explanation of Frank's truth-telling. So, it is incorrect to charge that "the pattern we need the attitudes to explain is only a pattern that is created by the attitude explanation." (Dretske, ms, 11)

I take it to be obvious that what one believes affects what one does. As I have argued in several places, I do not think that proponents of the Standard View have a satisfactory account of this fact. The only proponent of the Standard View who tells a remotely plausible story about the causal relevance of beliefs is Dretske himself.² But even Dretske admits that his view “doesn’t make what is believed and what is desired (i.e., content) relevant to what we do....” (Dretske, ms, 10) So, I do not think that the Standard View has a thing over Practical Realism in terms of showing how what one believes affects what one does.

Let me conclude this section by responding to two of Elugardo’s points (Elugardo, ms, 24ff) about my view of the causal explanatoriness of belief. In *Explaining Attitudes*, I gave three conditions for an explanation (pre-theoretically) to be a causal explanation. Elugardo offers a counterexample. He says that the following satisfies the three conditions, but is not a causal explanation: “The Elugardo Thanksgiving feast was held at [Ray Elugardo’s] house in 1996 because it was [his] turn to host it.” I’ll bite the bullet: This is a causal explanation—albeit not a very informative one. The reason that Elugardo thinks that this is not a causal explanation is that the “fact that it was [his] turn to host it does not causally trace the fact that the dinner was held at [his] house on Thanksgiving Day, 1996.” (Elugardo, ms, 27)

I have two comments: (a) The purported explanation does cite an event that is in the causal history of the dinner: its being Elugardo’s turn to host the dinner led to Elugardo’s belief that it was his turn to host the dinner; since Elugardo is a dutiful family member, his belief that it was his turn to host the dinner led to the dinner’s being at his

² See, Fred Dretske, *Explaining Behavior: Reasons in a World of Causes* (Cambridge MA: MIT/Bradford Books, 1988). I criticized his account in “Dretske on the Explanatory Role of Belief,” *Philosophical Studies* 63 (1991): 99-112, to which Dretske replied in “How Beliefs Explain: Reply to Baker,” *Philosophical Studies* 63 (1991): 113-117. See also my response to Dretske’s reply in *Explaining Attitudes*, pp. 57-62.

house. So, in fact, we can trace a causal path from its being Elugardo's turn to host the dinner to the dinner's being at his house. (b) Whether or not a causal explanation allows us to trace a casual path through space and time from *explanans* to *explanandum* is irrelevant anyway. Think of social, political and economic explanations; we can know *that* one thing makes a difference to another without any idea about *how* it makes a difference. For example, you can know that the gnostic sayings of Allen Greenspan, Chairman of the Federal Reserve, make a difference to your prospects for a job in business without being able to trace a causal path from Greenspan's mouth to the person who hires you.

Elugardo has another counterexample—this one directed toward the Control Thesis. Here is the Control Thesis:

An occurrence of F in context C causally explains an occurrence of G if: (i) if an F had not occurred in C, then a G would not have occurred in C; and (ii) given that F did occur in C, an occurrence of G was inevitable. (EA, 122)

Here is the counterexample in full:

Imagine that my brother and I live in the same apartment. He is blind but can hear; I am deaf but I am sighted. To help us know when someone is at the door, the apartment is electrically wired in such a way that our doorbell will ring when and only when the light bulb in our doorway, which is always on, glows dim and bright in an alternating pattern. Because of the internal circuitry, the correlation between the doorbell ring and the light bulb glowing holds as a matter of law. Consequently, you cannot cause the light bulb to alternate between dim light and bright light by pressing the doorbell without also bringing about a situation in which my brother's eardrums vibrate (when he is home, within earshot, etc.). For, as a matter of law, doing the first is nomically correlated in this context with causing the doorbell to ring, which in turn normally causes my brother's eardrums

to vibrate. We may suppose that the doorbell ringing is the only cause of the vibrations in this context. Therefore, given the facts of the case, if the light bulb had not dimmed and then glowed brightly, my brother's eardrums would not have vibrated. And, given that the light bulb did glow in this pattern, the vibration of my brother's eardrums was inevitable. (Elugardo, ms, 29)

According to the Control Thesis, then, the light's dimming and brightening causally explains the vibrations in the ear of the blind person who can hear. But, Elugardo says, "causing the light to alternate between dimness and brightness does not causally explain why my brother's eardrums vibrate in that context." (Elugardo, ms, 29)

Given the context-dependence of causal explanation and the peculiarities of the context with the odd wiring, it is not at all obvious that we should disallow the lights alternating between dimness and brightness in a causal explanation of the vibrating in the brother's ears. The only reason that Elugardo offers for denying that the alternating dimness and brightness causally explains the brother's eardrums' is that "there is no causal path that traces the second kind of event [eardrum's vibrating] back to the first [alternating dimness and brightness] without bypassing the causal connection between the sound of the doorbell ringing and my brother's eardrums vibrating." But, is this so? The eardrums' vibrating is caused by the doorbell's ringing, which is caused by (nominally connected to) the alternating dimming and brightening of the light, which is caused by the pressing of the doorbell.

Since we are taking the peculiarities of the wiring to be part of the context (and thus fixed), any event along this path is a causal explanation of the eardrums' vibrating in this context. We have here a predictive/explanatory pattern (i.e., a counterfactual-supporting pattern). As Dennett has said, when one finds a certain kind of predictive pattern, "one has ipso facto discovered a causal power—a difference in the world that makes a

subsequent difference testable by standard empirical methods of variable manipulation.”³
Instead of showing that there is something wrong with the Control Thesis, Elugardo has furnished an unusual case of causal explanation.⁴

(2) The “Brain-Explain” Thesis and Metaphysical Constraints on Explanation

(Antony, Elugardo)

Perhaps the most persuasive argument for the Standard View is the argument from causal explanation, which I argued is unsound. Its first premise is the ‘brain-explain’ thesis:

IIIa. Unless beliefs were brain states they could not causally explain behavior.

I attack IIIa in part by means of a thought experiment: If it had turned out that our world was an Aristotelian world, then beliefs would not be brain states. Yet we would have the same causal explanations; beliefs would causally explain behavior in exactly the same way that they do in our world. So, the causal explanatoriness of beliefs does not depend upon their being brain states. Both Antony and Elugardo take issue with my argument. Consider Antony first.

Antony distinguishes between a strong and a weak reading of IIIa, and says that the thought experiment does not refute IIIa on the weak reading.. The weak reading is this:

IIIa-weak: In any possible world in which beliefs are explanatory, beliefs are physically realized internal states, and in this world those states are brain states. (Antony, ms, 11)

³ Daniel Dennett, “Real Patterns,” *Brainchildren: Essays on Designing Minds* (Cambridge MA: MIT/Bradford, 1998): 112.

⁴ For more on causal explanation, see Chs. 4 and 5 of EA, and my “Metaphysics and Mental Causation” in *Mental Causation*, John Heil and Alfred Mele, eds. (Oxford: Clarendon Press, 1993): 75-95.

It is true that I did not distinguish strong and weak versions of IIIa, but the thought experiment also works to refute IIIa-weak. Suppose that the world had turned out to be an Aristotelian world. (If, as I argued earlier, the Standard View is not necessarily true, then there are no grounds for disallowing that an Aristotelian view is merely possible.) In an Aristotelian world, there is no more reason to think that each instance of each belief is realized by a particular token of a physical internal state than there is to think that each instance of each belief is realized by a particular token of a brain state. It is not as though, had Aristotle been correct, each instance of each belief would be realized, not by a brain state token, but by a token of some other physical-internal state. Aristotle's view is not one that matches up particular instances of mental states with particular tokens of internal physical states, whether the internal physical states are brain states or something else. So, construing 'physical realization' in a way that distinguishes Antony's view from mine, she is just mistaken when she says that the thought experiment about the Aristotelian world "does not show that beliefs could be explanatory even if they failed to be physically realized *at all*." (Antony, ms, 11-12)

Antony reformulates the argument that has IIIa as a premise and replaces IIIa with

(1) Unless beliefs were *realized by some kind of physical states*, they could not causally explain behavior. (Antony, ms, 12)

With respect to this reformulated argument she has, "It is clear that the conceivability of an Aristotelian world has no bearing whatsoever on premise (1), which replaces Baker's (IIIa)." (Antony, ms, 13) But I was not just proposing that an Aristotelian world is conceivable. Assuming that an Aristotelian world is conceivable, my claim was that had an Aristotelian world been actual, beliefs would still be explanatory without being brain states. From an Aristotelian point of view, beliefs are no

more physically realized internal states that are not brain states than they are brain states. So, Antony's reformulated argument is as unsound as the original. For the Aristotelian thought experiment applies to the first premise of her reformulated argument—just as it does to IIIa.

Referring to the argument against IIIa, Antony says that “the argument does not show that brain states do not *constitute* or *realize* beliefs in the actual world.” Of course, it doesn't. But the argument against IIIa was not supposed to show that brain states do not constitute or realize beliefs in the actual world. IIIa is a premise in an argument for the Standard View, the argument from causal explanation. My rebuttal of IIIa shows that that particularly tenacious argument for the Standard View is unsound. Nothing more was intended.

Let me elaborate by looking at Antony's example of the explanation of the water's boiling in terms of sufficient heat. (Antony, ms, 15-16) She agrees that in an Aristotelian world, an explanation of the boiling of (the stuff that is phenomenologically just like) water in terms of sufficient heat would still be justified. And it “is up to further investigation to determine what precisely are the mechanisms, if any, that sustain the regularity.” But, she continues, “none of this has any bearing on the claim that in *this* world, water is H₂O.” Right. But, again, that's not the point. The immediate point is, as Antony has just conceded, that the explanation of the boiling in terms of the heat does not require that (the stuff phenomenologically just like) water be H₂O. I am here trying to reveal as unsound the argument for the Standard View that has as a premise the claim that in order to be explanatory, beliefs must be brain states. Refutation of the Argument from Causal Explanation does not show that beliefs are not brain states; it only shows that one central and tenacious argument for the view that beliefs are brain states is unsound.

Now turn to Elugardo's criticisms of IIIa.--the "brain-explain" thesis. From now on, I shall use Elugardo's label, '(BET)', for IIIa. So,

(BET) [=IIIa.] "...[U]nless beliefs were (either identical to or constituted by) brain states, they could not causally explain behavior."

Elugardo objects to the thought experiment about the Aristotelian world. He says that, proponents of the Standard View understand causal explanations in the sense of Aristotle's efficient causation; but if our world had turned out to be Aristotelian, then our belief-explanations would not be causal in the Aristotelian world in the sense of efficient causation. To this, I reply: I was not speaking of Aristotle's view on causation, but ours (whatever that is). If we do not accept Aristotle's four causes, we need not settle on which of the four causes belief explanations invoke. My point was that if we lived in an Aristotelian world in which beliefs are not brain states, then our belief explanations would be causally explanatory to the same degree and in whatever sense that they are now. My argument about the sense in which they are causally explanatory—namely, that they reliably satisfy our explanatory and predictive interests by supporting relevant counterfactuals—is not Aristotelian.

In addition to my argument based on the thought experiment about the Aristotelian world, I argued against (BET) by showing that the general conception of causal explanation that underwrites it is untenable. On Elugardo's recounting, I argued against (BET) on the grounds that it entails a false thesis, (RT):

(RT) All belief explanations are replaceable by brain-state explanations of the same phenomena.

The claim that (BET) entails (RT) can be restated as what Elugardo calls 'Baker's Premise:' "If beliefs could not causally explain behavior unless they were either identical

with or constituted by brain states, then belief explanations are replaceable by brain-state explanations of the same phenomena.” (Elugardo, ms, 17)

Elugardo has an argument against Baker’s Premise. He offers a counterexample that just assumes a view that I argued at length was false—namely, the view that an explanation of the destruction of a store by the fact that the building had been set on fire “metaphysically depends for its causal explanatory power on a deeper explanation that appeals to molecular facts about fires and buildings, soluble goods, etc., and their causal relations.” Since in *Explaining Attitudes*, Chs. 4 and 5, I argued against this approach to causal explanatoriness, I would reject Elugardo’s argument here as question-begging. Nevertheless, Elugardo and others (e.g., Antony) have raised enough doubts about “Baker’s Premise” that if I were writing this section today, I would avoid the premise and stick to the argument based on the thought experiment about an Aristotelian world and the argument from the general character of causal explanation.

One reason that defenders of the Standard View are so ardent about the “brain-explain” thesis is that, either implicitly or explicitly, they impose metaphysical constraints on causal explanations. (cf. Beckermann, Meyering) My rejection of such metaphysical constraints in favor of a more pragmatic approach marks a deep methodological divide between Standard Viewers of any stripe and me. For example, consider Antony’s relatively weak version of the Standard View.

Antony and I agree that we do not need to know anything about underlying mechanisms in order to have causal explanations. For example, the highway patrol officer asks you why you were speeding, and you give the true answer that you are late for a meeting. At this point, Antony and I part ways. For Antony takes it to be a metaphysical requirement that “in order to be a genuine cause of behavior a belief must be (either identical to or realized in) a brain state.” (Antony, ms, 15) No doubt there are

many underlying mechanisms at work in your brain, but there is no reason to suppose that any particular state of any underlying neural mechanism can be identified with this instance of a belief that you are late for a meeting. The apparent need to impose metaphysical preconditions on belief-explanations stems from acceptance of a comprehensive metaphysical picture—a picture that, in my opinion, is unwarranted. In contrast to those who accept that metaphysical picture, I see no reason to posit metaphysical preconditions that—if the Empirical Conjecture is correct—are undetectable. (And in making that point I am not misinterpreting Antony’s Non-Emergent Non-Reductive Materialism.)

So, I am in full agreement with Antony’s insistence that “[*k*]nowledge of the metaphysical preconditions for constituting a cause is never a prerequisite for finding acceptable a proffered causal explanation.” (Antony, ms, 14) I also concur that the epistemology of explanation, on which we agree, “is perfectly consistent with the metaphysical thesis that in order to be a genuine cause of behavior a belief must be (either identical to or realized in) a brain state.” (Antony, ms, 15) But consistency is not the issue in the debate over the Standard View. The issue is whether or not we are justified in endorsing some version of ‘beliefs as brain states’ as a particular metaphysical precondition on belief-explanations. Since all versions of ‘beliefs are brain states’ are subject to empirical (dis)confirmation--and if the Empirical Conjecture is correct, they will be be disconfirmed--it seems risky to me (but not to Antony) to push any version of this doctrine as a metaphysical constraint on belief-explanations.

(3) The Causal Powers of Beliefs (Meyering)

Much, if not most, of the disagreement between Meyering and me stems from our different conceptions of causation and causal explanation. There are two main areas of disagreement: (A) On my view, explanatory properties do not need to be locally

supervenient on the microproperties of their bearers; i.e., relational properties can be causally explanatory. (B) On my view, a macroexplanation does not need to be buttressed by a microexplanation to be successful.

(A) Professor Meyering provides a forceful argument that causal explanations invoke causal powers that supervene on local microstructure. He quotes McGinn as saying: “[T]he causal powers of a state must be intrinsically grounded; they cannot depend essentially upon relations to what lies quite elsewhere.” Then, Meyering asks pointedly, “How does Baker propose to address this issue?” (Meyering, ms, 3)

Like this: It is just false that causal explanations must invoke causal powers that supervene on local microstructure. If it were true, then much of what we care to explain would fall outside the purview of causal explanation altogether. The kinds of causal explanations of behavior that we actually have typically do not invoke properties that are intrinsically grounded, or that supervene on local microstructure. This is obvious when we look at the *explananda* of interest. The behavior that we want to explain is almost always individuated relationally: Why did the judge dismiss Paula Jones’s lawsuit against President Clinton? Why did Professor X vote to hire a person of marginal competence? -Why did John buy a ticket to Rio? No intrinsic properties of the judge or of Professor X or of John could possibly bring about the dismissal of a lawsuit, the vote to hire somebody, or the buying of a ticket to Rio. So, we know in advance that either (a) the properties invoked by causal explanations of ordinary behavior will not supervene on local microstructure, or (b) there are no causal explanations of ordinary behavior.

It is clear that Professor Meyering and I have different basic conceptions of causal explanation and causation. Although he quotes with approval my view that “theory and practice must be brought into reflective equilibrium,” he seems to think that there is a tension between that view and my rejection of the proposed a priori metaphysical

standards of explanatory adequacy. I see no tension here. Taking local supervenience as a constraint on causal explanations would simply rule out the possibility of there being causal explanations of ordinary behavior at all. This is so, once again, because ordinary behavior is itself individuated relationally. So, there is no prospect of reflective equilibrium between taking local supervenience as a metaphysical constraint on causal explanations and any practices of causally explaining ordinary behavior. Therefore, rejection of the local-supervenience constraint is not rejection of reflective equilibrium between theory and practice. Rather, acceptance of the local-supervenience constraint is rejection of causal explanations of behavior in the terms that interest us.

Notice that my line of argument depends on how *explananda* are individuated: Ordinary behavior, for which we want causal explanations, is not individuated in terms of intrinsic properties of the agent. It is not surprising that properties that explain ordinary behavior are not individuated narrowly when what they explain is not individuated narrowly. This line of argument is somewhat different from the familiar arguments about Twin Earth cases. So, I am not relying on the assumption that “just any belief sentence is susceptible to Twin Earth examples.” (Meyering, ms., 7)

Meyering distinguishes between “the problem of here-and-now causal efficacy, as opposed to that of causal relevance in some wide sense.” What he calls ‘the problem of causal efficacy’ concerns “the question which properties of my internal states here-and-now are causally responsible for making my arm move, or whatever. And an answer to this question must surely advert to local properties of my internal states.” (Meyering, ms, 7) If the question were, which properties here-and-now bring it about that your arm moves, then I would agree with Meyering that we advert to local properties of my internal states. But that question has almost nothing to do with our concerns to explain behavior. We rarely want to know why your arm moved; we want to know why you

voted a certain way—no matter how the vote was taken, whether by raising your hand, or writing on a ballot, or something else. Focusing on the local properties of the internal states that are causally responsible for making your arm move will never yield a causal explanation of your vote. If your vote has any causal explanation, it is not in terms of properties that supervene on your local microstructure. So, to insist on the local-supervenience constraint for causal explanation is to give up on causal explanation of ordinary behavior altogether.

To call rejection of a local-supervenience constraint “action-at-a-distance obscurantism” (Meyering, ms, 5) is simply to use loaded language. And this loaded language is unjustified in light of the fact that the local-supervenience constraint on causal explanation itself has the consequence that ordinary behavior would have no causal explanation at all. Moreover, we all accept explanations that Meyering would describe as “action at a distance.” For example, why didn’t B, who was the most qualified applicant, get the job? The explanation is that A started rumors about B, and the potential employer, who believed the rumors, decided not to hire B. Whether or not this is a good causal explanation has nothing to do with whether or not there is a sequence of events beginning with events in A’s body and ending with some event in B’s body. A can injure B without ever having any physical connection with B.

We no more need narrow psychological explanations (explanations on the basis of intrinsic properties of the individuals whose behavior is to be explained) than we need narrow economic explanations. No one demands narrow economic causation: we can explain why Jones’s money went to the ABC pension fund by citing the facts that Jones is a state employee and that payments to the ABC pension fund are automatically taken out of state employees’ paychecks. Although the property of being a state employee—which causally explains the diversion of Jones’s money to the ABC pension fund—is not

an intrinsic property of anybody, no one is tempted to think of economic explanations as tainted. Relational psychological properties are no more suspect than are relational economic properties.

(B) Meyering, like many other Standard Viewers, believes that macroexplanations “cannot just stand on their own.” (Meyering, ms, 10) Although Meyering does not reject the explanatory legitimacy of macroexplanations, he says: “Whenever macroexplanations are being offered, there is always a warranted presumption that in principle a more detailed account can be given of the causal powers of the macro-properties invoked in terms of some relevant implementing mechanisms operative at the next level down in some more or less determinate functional hierarchy of sciences and their respective domains.” (Meyering, ms, 10) What Meyering describes here is a very substantial metaphysical view that I do not share. My view is this: Domains of different sciences may or may not be nested in the way that Meyering’s view requires. Reduction is where you find it. You have to look and see. There are no guarantees in advance about the unity of science. I do have a further conjecture: If the sciences do turn out to be unified in the way that Meyering suggests, then many things for which we want causal explanations will not be explainable by any of the sciences.

Since Meyering says that I endorse a thesis of “*radical* autonomy for intentional psychology,” let me say again what my view is. I do not hold that any old type of intentional explanation is impervious to the findings of neuroscience. However, I do believe that no future discovery of neuroscience could force us to abandon our global intentional explanatory frameworks. The basis for my belief is not the theological caricature that Meyering proposes at the end of his paper, but is the fact that we—scientists and nonscientists alike—are enmeshed in a comprehensive commonsensical framework, without which science itself would be impossible. Since I give arguments for

my outlook both in *Saving Belief* and in *Explaining Attitudes*, I will not rehearse them here.

(4) Microreduction (Beckermann)

Beckermann appeals to considerations of the generality of microphysics in defense of the Standard View, as he construes it. He construes the Standard View as the following thesis: Mental properties are either identical with or reducible to physical properties. This characterization of the Standard View is much stronger than mine. Many philosophers who are proponents of the Standard View in my sense—e.g., functionalists, philosophers in Jerry Fodor’s camp—would deny that mental properties are either identical with or reducible to physical properties. For example, Jerry Fodor gives a spirited argument against reductionism in “Special Sciences.”⁵ Fodor holds that every instance of a mental property is an instance of some physical property or other; but he denies that for every mental property M, there is a physical property P such that every instance of M is an instance of P. That is, Fodor denies that mental properties are either identical with or reducible to (in a familiar and intuitive sense) physical properties. And of the Standard-View contributors to this volume, Antony, Dretske and Elugardo would reject both identity and reduction of mental properties to physical properties.

So, my construal of the Standard View is more encompassing than is Beckermann’s. But what is important are the positions in question, not which ones that we assign a certain label like ‘Standard View.’ Since Beckermann argues for a reductionist view, he would reject the non-reductive view that belief-tokens are merely constituted by neural tokens, no matter what the view is called. Since I oppose both non-reductive and reductive versions of ‘beliefs are brain states,’ I continue to prefer my broader characterization of the Standard View. Nevertheless, Beckermann’s construal of

⁵ Jerry Fodor, “Special Sciences” in *Representations* (Cambridge MA: MIT/Bradford, 1981): 127-145.

the Standard View is at least part of what I consider the Standard View. So, I want to discuss Beckermann's thesis that I shall call '(BC),' short for 'Beckermann's Construal of the Standard View:'

(BC) Mental properties are either identical with or reducible to physical properties.

Beckermann gives a careful account of (BC), by defining 'identity' and 'reducibility.' He begins with C.D. Broad's distinction between mechanically explainable properties and emergent properties, which he summarizes in two definitions, (ME) and (E)). This "distinction between mechanically explainable and emergent properties," says Beckermann, "far better captures the intuitive difference between reducible properties and properties that cannot be reduced to more fundamental properties much better than all other accounts have been able to...." (Beckermann, ms, 8)

I do not think that Broad's distinction between mechanically explainable and emergent properties does capture the intuitive difference between reducible and irreducible properties. For one thing, the reducible/irreducible distinction is exhaustive (every property is one or the other); but the mechanically-explainable/emergent distinction is not exhaustive. To see this, note that (ME) and (E) have the same first clause, (a): "The statement 'For all x: if x has the micro-structure $[C_1, \dots, C_n; R]$, then x has the macro-property F' is a true law of nature," where x has the micro-structural property $[C_1, \dots, C_n; R]$ if and only if x consists of the parts C_1, \dots, C_n which stand in the (spatial) relation R to each other. (Beckermann, ms, fn6) So, if there is a system that has macro-property F, and that consists of the parts C_1, \dots, C_n standing in the (spatial) relation R to each other, and if there is no law of nature of the relevant form, then the property F is neither mechanically explainable nor emergent.

Now consider the property of being married. There is no true law of nature that says: “For all x: if x has the micro-structural property $[C_1, \dots, C_n, R]$, then x has the macroproperty of being married.” Therefore, the property of being married is neither mechanically explainable nor emergent, according the definitions (ME) and (E). But the property of being married, like every other property, is either reducible or irreducible. So, the intuitive difference between reducible and irreducible properties is not adequately reflected in the distinction between mechanically explainable and emergent properties.

On Beckermann’s account of reducibility, (R), a condition for macro-property F of a system S at t to be reducible to a micro-structural property $[C_1, \dots, C_n, R]$ is this: “it follows from the general laws of nature applying to the components C_1, \dots, C_n , that S at t possesses all features which are characteristic of property F.” (Beckermann, ms, 8) Since it does not follow from the laws of nature that anyone with components C_1, \dots, C_n in a certain spatial arrangement is married, the property of being married is not reducible. to micro-structural properties.

Now consider the mental property of believing that some people are married. According to (BC), the mental property of believing that some people are married is identical with or reducible to a physical property, in Beckermann’s sense of reducibility, (R). On (R), if S’s believing that some people are married is reducible, then “it follows from the general laws of nature applying to the components C_1, \dots, C_n that S at t possesses all features which are characteristic of the property” of believing that some people are married.

It seems implausible to suppose that S’s having the property of believing that some people are married follows from general laws of nature applying to S’s microstructure. For a necessary condition for S to have the property of believing that some people are married is that S have the concept of *being married*. But S can have the

concept of being married only in certain social environments, where there are conventions and civil laws. (It's not on account of the fundamental properties of the particles that make up a dog's brain, say, that a dog lacks the property of believing that some people are married.)

So, I do not think that S's believing that some people are married follows from general laws of nature applying to S's components. If that is right, then the property of believing that some people are married is a mental property that is not reducible (in Beckermann's sense) to physical properties. And if it is not so reducible, it is not identical to physical properties either. And if there is a mental property that is neither identical with nor reducible to physical properties, then (BC) is false.

In sum, Beckermann construes the Standard View to be the thesis (BC): all mental properties are identical with or reducible to physical properties, where he understands 'physical properties' nonrelationally. The property of believing that some people are married is a mental property that is neither identical with nor reducible to physical properties in the sense of Beckermann's (R). So, here is a counterexample that falsifies the Standard View as Beckermann construes it. Now let us turn to Beckermann's general metaphysical argument in support of the Standard View as he construes it.

Beckermann calls his general metaphysical argument for (his version of) the Standard View the 'argument from the implausibility of downward causation.' Here is a reconstruction of the argument:

- (1) If (BC) were false, then "mental properties would at least be emergent, i.e., properties that, even in principle could not be deduced from the parts of the organisms that are their bearers." (Beckermann, ms, 11)

(2) If any properties are emergent, then there is a gap in nuclear physics: “atoms are sometimes moved by forces the existence of which cannot be derived from the general laws of nuclear physics.” (Beckermann, ms, 12)

(3) Such a gap is implausible.

Therefore,

(4) [We have good reason to think that] (BC) is true.

This argument raises very large metaphysical issues that I cannot fully address here. So, I shall focus rather narrowly on this argument (1) – (4). First, I shall comment on the use of the term ‘emergent.’ The term ‘emergent’ was defined by (E). (E) has two clauses. A property’s being such that “even in principle [it] could not be deduced from the parts of the organism[] that [is] its bearer” is only a necessary, not a sufficient condition for the property’s being emergent according to (E). If we interpret ‘emergent’ in (1) as it was defined by (E), then (1) is false. For if there is a mental property that is neither identical with nor reducible to physical properties, then (BC) is false. And the antecedent of (1) is true. But it does not follow that mental properties are emergent in the sense of (E). For the definitions of emergence and reduction—(E) and (R)—that Beckermann gives are not exhaustive: a property may fail to be *both* emergent and reducible. And the property of believing that some people are married is such a property. Earlier, I argued that the property of believing that some people are married is not reducible in the sense of (R); and it is not emergent in the sense of (E) for the same reason: There is no true law of the form: “For all x: if x has the micro-structure [C₁,...,C_n; R], then x has the macro-property of believing that some people are married.” So, on the definition of ‘emergence’ that Beckermann gives, it is incorrect to suppose that denial of (BC) commits one to the existence of emergent properties in the sense of (E). So, given (E), premise (1) is false.

However, in his general metaphysical argument, Beckermann regards emergent properties as “properties that, even in principle, could not be deduced from the properties of the parts of the organisms that are their bearers.” And I do hold that mental properties are not deducible from the (nonrelational) properties of the parts of the organisms that have the mental properties. Now according to (2), this view would have the consequence that there would be a gap in nuclear physics. The putative gap is that “atoms are sometimes moved by forces the existence of which cannot be derived from the general laws of nuclear physics.” I want to show that my view, according to which mental properties are not deducible from the (nonrelational) properties of the parts of the organisms that have the mental properties, does not entail that there is an implausible gap in nuclear physics.

Beckermann is correct to say that I claim that many properties are such that even in principle they could not be deduced from the micro-structural properties of the parts of their bearers. But it does not follow from this that such properties could not be deduced from *any* micro-structural properties. For example, it is consistent with my view that both (i) the property of believing that some people are married is not deducible from properties of the believer’s micro-structure, and (ii) the property of believing that some people are married is deducible from the properties of the believer’s micro-structure together with other micro-physical properties borne by other things in the believer’s social environment. In short, my view is compatible with what is called ‘global supervenience.’

Global supervenience is, roughly, the view that for any two possible worlds that are exactly alike in micro-physical laws and in distribution of micro-physical particles are exactly alike in every respect. Since my view is compatible with global supervenience, it does not entail that there is an implausible gap in nuclear physics. For if the thesis of

global supervenience is true, then there are no changes in macro-physical properties without changes in micro-physical properties.

Indeed, my view of mental properties is even compatible with Jaegwon Kim's thesis of strong supervenience, as long as the base properties need not be intrinsic properties of the bearer of the supervening property.⁶ If—as is logically possible—my view were combined with a relational form of strong supervenience, then “downward causation” could be avoided altogether. Putative macrocausation would be merely apparent. All genuinely causally efficacious properties would be properties of fundamental particles.⁷ I am not proposing that this is the case; I am only pointing out that my view of the attitudes does not entail “downward causation.”

In sum: I am not committing myself to any supervenience thesis; indeed, I do not think that we are in any epistemic position to affirm or deny any universal supervenience thesis. My point here is only that I have contravened neither the thesis of global supervenience nor the thesis of a relational form of strong supervenience. If Beckermann understands ‘general laws of nuclear physics’ in such a way that there still follows from my view a gap in the general laws of nuclear physics, I would not see such a “gap” as implausible at all.

(5) Non-Emergent Non-Reductive Materialism (Antony)

Antony helpfully provides a finer-grained taxonomy of views on the attitudes than I was using. However, I believe that my arguments work as well on the finer-grained taxonomy as they do on the taxonomy that I was using. Although I agree that Antony's taxonomy is more interesting than mine, the differences between our taxonomies are irrelevant to my arguments against the various versions of the Standard View, as I shall

⁶ See Jaegwon Kim, “Concepts of Supervenience” in *Supervenience and Mind: Selected Philosophical Essays* (Cambridge: Cambridge University Press, 1993): 53-78.

⁷ See Jaegwon Kim, “Epiphenomenal and Supervenient Causation” in *Supervenience and Mind: Selected Philosophical Essays* (Cambridge: Cambridge University Press, 1993): 92-108.

try to show. Antony identifies her view as Non-Emergent Non-Reductive Materialism and mine as Emergent Non-Reductive Materialism. (Frankly, at this point, I do not know in exactly what sense, if any, I would answer to the name ‘emergentist.’) Taxonomies aside, the main difference between Antony’s and my positions is this: Whereas Antony thinks that for each instance of each belief there is a token of a physical internal state that is identical to (or constitutes or realizes) that belief, I do not.

In this section, I just want to point out that Antony’s Non-Emergent Non-Reductive Materialism does not escape the net of my arguments against the Standard View. According to any version of the Standard View, particular instances of beliefs should be construed to be particular neural tokens. Antony states her position as follows: “Representationalists like me believe that the best account of how beliefs cause actions is one that posits mental representations: neural assemblages that are type-identified by their structural (or syntactic) properties, and which properties in turn mirror their semantic properties.” (Antony, ms, 21) If there are such “neural assemblages that are type-identified by their structural (or syntactic) properties, and which properties in turn mirror their semantic properties,” these neural assemblages should be discoverable by neuroscientists. (It seems to me idle to postulate neural states undetectable by neuroscientists.) My Empirical Conjecture is that neuroscientists will not discover neural assemblages with the requisite properties to be construed as instances of particular beliefs. Nothing known about the brain (by me, anyway) offers any hope that the brain is organized in such a way that for each instance of each belief there is a token brain state. If the Empirical Conjecture is correct, then no version of the Standard View, including Non-Emergent Non-Reductive Materialism, is warranted.⁸

⁸ Or, as I put it in ABBS, if the Empirical Conjecture is correct, then no version of the Standard View is true. Elugardo, Antony and others have taken me to task for confusing being unwarranted with being untrue. I can agree with the extreme realist that it is logically possible for a theory to be (forever) unwarranted but true. But granting such logical possibility does not prevent one from plausibly concluding that if a theory is (forever) unwarranted, it is untrue. Certainly, if a theory is (forever) unwarranted, then

Antony criticizes a dilemma that I posed for the Standard View. If any version of the Standard View is correct, then either beliefs supervene on brain states or beliefs are (merely) constituted by brain states.⁹ I argue that beliefs do not supervene on brain states. And I argue that if beliefs are merely constituted by brain states, then the Standard-View thesis is explanatorily uninteresting: “Even when there is an undisputed relation of constitution between an intentionally identified object and material elements that constitute it, the constituting elements shed no explanatory light on the constituted object as intentionally identified.” (EA, 183) So, the first horn is false (as Antony agrees without giving me credit for my argument), and the second horn renders the Standard View explanatorily uninteresting.

Antony (rightly) opts for the second horn and objects that “the thesis that beliefs are constituted by brain states has a great deal of explanatory import—it explains how beliefs can be causes.” (Antony, ms, 24) I have two comments: (i) Much of the recent literature on mental causation argues, successfully in my opinion, that even if beliefs are brain states that cause behavior, the fact that a certain behavior-causing brain state is a belief is explanatorily irrelevant.¹⁰ That is, the fact (if it is a fact) that a certain brain state is or constitutes a certain belief is irrelevant to what that brain state causes. It is not in virtue of being or constituting a belief that a brain state causes what it does. So, the Standard View, even if correct, would not show how beliefs qua beliefs can be causes. (ii) In any case, if brain states answering to a Standard-Viewer’s description cannot be

no one is justified in regarding it as true.

⁹ Antony’s term of choice is ‘are realized by.’ Although she uses the term unequivocally, it is often used equivocally in the literature to refer both to supervenience and constitution. So, I avoid the term ‘realization.’

¹⁰ The issue was raised in “Mind Matters,” by Ernest LePore and Barry Loewer, *Journal of Philosophy* 84 (1987): 630-642. In “Metaphysics and Mental Causation,” (in *Mental Causation*, John Heil and Alfred Mele, eds. (Oxford: Clarendon Press, 1993: 75-96), I tried to show that, given a certain metaphysical picture, the problem is intractable. As I’ve mentioned, Dretske is the only Standard-View philosopher I know who has really wrestled with this problem.

found, then the “explanation” of how beliefs can be causes is a mythical one. Appeal to Shoemaker’s distinction between core realizations and total realization is to no avail here. If the Empirical Conjecture is correct, then Shoemaker’s distinction between core realizations and total realization breaks down: the core realizations would be merely speculative items, like the ether (before Michelson-Morley). And, also like the ether, core realizations are dispensable. What would be relevant for understanding mental causation would be total realizations—which may include the entire brain, and from which it may not be possible to “factor out” core realizations.

(6) The Master Argument Against the Standard View (Dretske, Antony, Elugardo)

Dretske charges me with two mistakes in what he derisively calls “the ‘proof’”—the Master Argument against the Standard View: (a) a mistake of supposing that “a belief that *p* is always realized in (a token of) the same type of brain state,” (Dretske, ms, 14) and (b) a mistake of “thinking that according to standard theory, mental states are the proper study of *brain* science—i.e., neurophysiology.” (Dretske, ms, 13) Neither charge hits home.

(a) As to the first alleged mistake, I simply deny that I supposed that “a belief that *p* is always realized in (a token of) the same type of brain state.”¹¹ Rather I suppose this: If ‘beliefs are brain states’ is an empirical thesis, then the neural tokens that are tokens of a single belief (at least in an individual) must have some neurophysiological features in common. There are both metaphysical and epistemological considerations that support this claim.

Dretske’s vending-machine analogy suggests a metaphysical reason for my claim. A slug has the same causal role in the vending machine as a legitimate coin, and it has the

¹¹ In my characterization of the Standard View, (SV), by saying “(ii) *S* tokens *n*,” I meant that the token *n* occurs in *S*’s brain. After some labor, Dretske ends up interpreting (SV) in the intended way. I take my target to be exactly what Dretske insists it must be.

same effect as a legitimate coin: both slug and coin bring forth candy bars. If the machine “keys” only, say, shape and thickness of legitimate coins, then, in order to play the same causal role as the coin, the slug must have the same shape and thickness as a coin. But the slug may be very different from the coin in other physical respects (e.g., it may be made of a different material, have a different color, etc.). Hence, the slug need not be of the same physical type from the coin. Nevertheless, the slug must share with the coin the physical properties that enable it to play the same causal role in the vending machine as the coin. Similarly for token brain states that are (according to the Standard View) token beliefs. Two neural tokens of different physical types may be the same beliefs only they have some physical properties in common. It is in virtue of those physical properties that the two neural tokens have the same causal role.¹²

Let me make this point from a different angle. suppose that beliefs are brain states, as the Standard View claims, but that type-identity is false. My claim is that there still must be salient neurophysiological features in common with O belief tokens (at least within a single individual). Dretske asks rhetorically: “If O detectors need not have anything in common (besides occurring, or having the function of occurring, in the presence of Os), why must O beliefs have some salient *neurophysiological* feature in common?” (Dretske, ms, 15) Here is a straightforward answer:

If particular O detector had nothing in common with other O detectors (besides occurring, or having the function of occurring in the presence of Os), then, in order to be an O detector, it must have detected Os in the past.¹³ Otherwise, it would not have the

¹² I take this to be a consequence of functionalism. It is highly unlikely that two neural tokens of different types that a functionalist identifies with a single belief have exactly the same causal role since their causal roles will be affected by the subject’s other attitudes. This holism raises a difficulty for individuation of causal roles, and hence a difficulty for construing beliefs in terms of causal roles. But that’s another problem, not the one at issue here.

¹³ This is true only for natural, not artifactual, O detectors.

function of detecting Os. (This is roughly Dretske's view.¹⁴) If the analogy holds, then if an O belief token (=a neural token) had nothing in common with other O belief tokens (=neural tokens), then, in order to be an O belief, it must have detected Os in the past. But this is impossible: for a belief/neural token occurs only once; so *it* could not have detected Os in the past. Therefore, in order for a neural token to be an O belief token, it must have something in common with other neural tokens that detected Os in the past. And what the current neural token has in common with other neural tokens that detected Os in the past must be a neurophysiological feature. Otherwise, the fact that past neural tokens detected Os would have no bearing on whether or not the current neural token is an O belief or has the function of detecting Os. So, if the analogy between o detectors and O beliefs holds, then a neural token that is an O belief must have some salient neurophysiological feature in common with past tokens that detected Os.

Note that the claim that neural tokens of the same belief type have some neurophysiological feature in common does not entail type-identity of beliefs and brain states. For even if tokens of *different* neurophysiological types were O beliefs, there would have to be many tokens of *each* of the types that could be O beliefs. To be an O belief, or an O detector, or to have the function of occurring in the presence of Os, a neural token can not be a "one-shot" token that has no salient neurophysiological properties in common with other neural tokens that have detected Os in the past.

So, my claim that on the Standard View, there must be neurophysiological similarities among neural tokens that are tokens of a single belief is not a covert appeal to type-identity. For we have good metaphysical reason to hold that even if type-identity is false, instances of a particular type of belief cannot be neural tokens unless those neural

¹⁴ See Fred Dretske, *Explaining Behavior: Reasons in a World of Causes* (Cambridge MA: MIT/Bradford, 1988).

tokens share salient neurophysiological features. And we have good epistemological reason as well. So, turn now to an epistemological consideration.

The claim that belief tokens are neural tokens is supposed to be an empirical claim. There is no reason to accept any empirical claim in the absence of confirmation. The way to confirm the hypothesis that a particular neural token *n* is a token of an *O* belief, say, would be to compare the neural token *n* to other neural tokens that are candidates for being tokens of an *O* belief. If the neural tokens that are candidates for being tokens of an *O* belief had no neurophysiological feature in common, there would be no way to carry out the comparison: there would be no way to distinguish between a correct identification and an incorrect identification of a particular neural token with an instance of an *O* belief.¹⁵ A theory that postulates identification of tokens without any way to ascertain whether a given proposed identification is correct is not confirmable or disconfirmable. Hence, if neural tokens that are tokens of the same belief had nothing neurophysiologically in common, then the view that neural tokens are belief tokens would not be an empirical theory.

In sum, my point is different from the mistaken point (a) that Dretske attributes to me. For my point does not presuppose that the Standard View is committed to type-identity of beliefs and brain states. Rather, my point is that in order for a neural token to be a candidate to be a belief that *p*, it must have some salient neurophysiological features in common with other neural tokens that are candidates (at least in a single brain) to be a belief that *p*.

(b) The second “mistake” that Dretske charges me with is to think that, “according to standard theory, mental states are the property study of *brain* science—i.e.,

¹⁵ This point also speaks to Elugardo’s analogy of the Gothic statue. In the case of Gothic statues, we can identify pieces of marble as candidates to be Gothic statues and then investigate whether each is a Gothic statue. But in the case of neural tokens with no neurophysiological similarity, there is no way to distinguish those neural tokens that are even candidates to be instances of a belief that *p*.

neurophysiology. “Why think neuroscience is the relevant science to identify beliefs?” he asks. (Dretske, ms, 14) Dretske thinks that the only reason that one would look to neuroscience is that one is covertly assuming type-identity. I have just shown that I do not covertly assume type-identity. There are other reasons for my claim that neuroscience is the relevant science for confirming any version of the Standard View.

The obvious reason is that the Standard View *is* the view that beliefs, if there are any, are brain states; and neuroscience is the science that studies brain states. Unless there are many neural tokens that can be identified as instances of various beliefs, there is no confirmation of the Standard View. And the only science in a position to identify and classify neural tokens is neuroscience.

This conclusion is reinforced when we consider the alternatives to neuroscience that Dretske suggests. Dretske asks: “Why shouldn’t evolutionary biology or the learning history of the organism be the relevant source to consult in determining whether any particular state of the creature is a belief that *p*?” (Dretske, ms, 14) This rhetorical question invites us to ignore the problem that I am pointing to. For he is assuming that we already have picked out a particular state (token) of a type whose evolutionary or learning history can be checked. But in order to determine whether a particular state (token) of a creature is a belief that *p*, there must be some way to pick out the state (token)—other than as an instance of the belief that *p*. In order to see whether or not a particular state (token) is of a type that has a particular evolutionary or learning history, we must be able to pick out the particular state (token) in question. So, the picking out of a particular state (token) of a creature in order to determine whether it is a belief that *p* is logically and temporally prior to determining whether that particular state is a token of a type that has a certain evolutionary or learning history. For this reason, neither evolutionary biology nor the learning history of the organism is of help in picking out the

particular state (token) in question. So, we are back to neuroscience as the science equipped to pick out current token states that are plausible candidates to be instances of belief.

I have no doubt that there is an evolutionary-biological story about how brains enable us to have beliefs. But that story does not address the problem that I am raising. What I'm calling the Standard View concerns identification of particular neural tokens as instances of various beliefs. This requires that relevant current tokens be picked out in order to determine (perhaps by consulting the evolutionary history of tokens of that neural type) that they are instances of the belief that p. From the point of view of the Standard View, appeal to evolutionary biology or learning history might supplement neuroscience; but neither could replace it without abandoning the Standard View.

Both Dretske and (implicitly) Antony complain about premise 2.221 of the Master Argument in "Are Beliefs Brain States?" Premise 2.221 concerns noneliminative theories according to which particular beliefs are particular brain states. According to 2.221, "If T is necessarily true, then the correct account of belief is either a form of species chauvinism or uses the term 'belief' equivocally"—where T be any noneliminative theory according to which particular beliefs are particular brain states. The reasoning supporting 2.221 is that if T is necessarily true, then T reveals the nature of belief; and if T gives the nature of belief, then there is no belief that fails to be a brain state—in which case nothing that fails to have a brain (a computer or a god) could possibly have a belief. Both Dretske and Antony construe the Standard View more broadly than 2.221 allows. Antony in effect says that we should we take the relevant necessary truth to be not, that beliefs are brain states, but, more comprehensively, that beliefs are realized by some kind of physical states.¹⁶

¹⁶ Antony is here attacking my argument from causal explanation in *Explaining Attitudes* (Antony, 16 ms.). I am discussing this point in the context of 2.221 on the basis of a referee's criticism that 2.221 is

I take the criticism to amount to this: If T is re-formulated as T'—a theory about physically-realized states, rather than about brain states—then 2.221 is false: A theory about physically-realized states may be necessarily true without being a form of species chauvinism and without using any terms equivocally. So, let us reformulate the argument for 2.22, letting T' be any noneliminative theory according to which particular beliefs are particular physically-realized states:

2.221* If T' is necessarily true, then it is metaphysically impossible that
Cartesian dualism be true.

2.222* It is not metaphysically impossible possible that Cartesian dualism be
true.

∴ 2.22 T' is not necessarily true.

Premise 2.221* is, I think, uncontroversial. If a theory is metaphysically impossible, then there is no possible world in which it is true; if a theory is necessarily true, then it is true in every possible world. If T' is necessarily true, then it is metaphysically impossible that there be any particular belief that fails to be a particular physically-realized state; if it is metaphysically impossible that there be any particular belief that fails to be a particular physically-realized state, then Cartesian dualism—according to which there are beliefs that are states of an immaterial soul and hence not physically-realized—is metaphysically impossible. So, 2.221* is true.

The controversial premise in the reformulated argument is 2.222*. I have no direct argument for the metaphysical possibility of Cartesian dualism, but I do have an indirect consideration. Along with many Standard Viewers, I have a strong intuition that Cartesian dualism is metaphysically possible (though not that it is true). Perhaps some

false.

Standard Viewers think that that intuition is wrong—that Cartesian dualism is not even metaphysically possible. How do we adjudicate?

The only argument that I can think of that would support the metaphysical impossibility of Cartesian dualism is a Kripke-style argument that shows what would lead people to endorse the intuition of the metaphysical possibility of Cartesian dualism and yet be wrong. On a Kripke-style argument,¹⁷ we confuse the metaphysical intuition about the possibility of Cartesian dualism with a similar, but distinct, epistemic intuition. The epistemic intuition is that we could be in an epistemic situation indiscernible from the one that we actually are in, but one in which Cartesian dualism is metaphysically impossible. It is the epistemic intuition, not the metaphysical intuition—the argument would go—that is correct. On this approach, those who endorse the metaphysical possibility of Cartesian dualism are confusing that metaphysical intuition with the related epistemic intuition. This approach depends on reconstruing the (false) metaphysical intuition about what is metaphysically possible as a (true) epistemic intuition about what is compatible with the evidence that we have.

But this argument against the intuition that Cartesian dualism is metaphysically possible would be a misapplication of a Kripke-style argument.¹⁸ In a Kripke-style argument, what motivates reconstrual of the intuition about metaphysical possibility is the fact that we have *independent reason* to think that the original intuition about metaphysical possibility is false. *Given* that the original intuition about metaphysical possibility is false, we cast about to find why the original intuition seemed so compelling. In Kripke-style argument, what we find is that there is an epistemic intuition (i) that is true and (ii) that we confuse with the original metaphysical intuition. Once we see this, we replace the original metaphysical intuition with the correct epistemic intuition. But

¹⁷ I am not attributing any such argument to Kripke himself.

¹⁸ For a discussion of a different (mis)application of a Kripke-style argument, see my “Why Constitution is Not Identity” (*Journal of Philosophy* 94 (1997): 599-621, especially 615-619).

the original intuition about the metaphysical possibility of Cartesian dualism not like this at all. We have no independent reason to suppose that the original intuition about the metaphysical possibility of Cartesian dualism is false in the first place.

To see the disanalogy between an intuition that a Kripke-style argument does discredit and an intuition—such as the intuition that Cartesian dualism is metaphysically possible—that a Kripke-style argument does not discredit, consider the (incorrect) metaphysical intuition that heat is not necessarily molecular motion. Given Kripke's essentialism, the discovery that heat is molecular motion disconfirms the original intuition that heat is not necessarily molecular motion. It is that disconfirmation that motivates the reconstrual of the metaphysical intuition as an epistemic intuition. In the case of the Standard View, there has been no parallel *discovery* that beliefs are brain states (or more generally, are physically-realized states). And as I have argued throughout, I take it to be an empirical conjecture that there will be no such discovery. Thus, there is no justification for discarding the original metaphysical intuition that Cartesian dualism is metaphysically possible. The only reason to suppose that Cartesian dualism is metaphysically impossible is a prior commitment to the necessity of the Standard View. In the current context, such a commitment would be obviously question-begging.

Since the only argument that I can think of against the metaphysical possibility of Cartesian dualism is flawed (and since almost all philosophers since the seventeenth century have believed in the metaphysical possibility of Cartesian dualism), I think that it is reasonable to suppose that the intuition of the metaphysical possibility of Cartesian dualism is correct. But if Cartesian dualism is even metaphysically possible, then it is metaphysically possible that there be a particular belief that fails to have a particular physical realization. In that case, 2.222' is true, and the argument for 2.22 is sound.

Elugardo has several epistemological objections to the Master Argument. I shall respond to two of them. First, Elugardo points out there is confirmation of the view that genes are constituted by stretches of DNA even though there is “no neurochemical property that they all must have in common in virtue of which they are a particular type of gene.” (Elugardo, ms, 10) True enough, but there are complications in the belief case that have no parallel in the DNA case. I doubt that even *within* a given individual S, there is a single type of neurophysiological property that is instantiated on every occasion on which S does something explainable by S’s belief that p. To cite an example from “Are Beliefs Brain States?”: Suppose that Jones wants to improve his social status, and believes that the best way to improve his social status is to become well-known in the community. He acts on this belief in various ways: he makes large contributions to charities, becomes active in community organizations, and so on. But (in line with my Empirical Conjecture) I doubt that there is some particular neurophysiological property that is instantiated on each occasion of his acting on the belief that the best way to improve his social status is to become well-known in the community.

Second, Elugardo objects to my use of the term ‘confirmation’ in premise 2.231 (“If T is contingently true, then T will be confirmed by neuroscience.”) I meant to use ‘confirmation’ univocally, in the sense that Elugardo calls ‘experimental confirmation:’ “[T]o confirm T is to have positive instances of its generalization[s].” (Elugardo, ms, 9) I agree with Elugardo that it is possible that there be a theory that is true, empirical and yet, in fact, never experimentally confirmed. But, barring a Cataclysmic Event, I think that in the case of T, it is unreasonable to suppose that there will be such a theory that is true, empirical and yet, in fact, never experimentally confirmed. Absence of confirmation in the long run would be strong evidence that there is no true T.

(7) Practical Realism Extended (Meijers)

Professor Meijers focuses on an important but neglected class of attitudes: collective beliefs. Consider, for example, “The Security Council decided to send UN troops to Albania because it believed that it was necessary to restore law and order.” Meijers convincingly shows that a collective belief, such as that of the Security Council, cannot be understood either as an aggregate of individual members’ beliefs or as attitudes of some kind of “super-agent” that is independent of individuals. Meijers has done a real service by drawing attention to such beliefs, which, I confess, I had never thought about at all.

Meijers has two aims: First, to compare the Standard View and Practical Realism with respect to how well each can handle collective beliefs; and, second, to use collective beliefs to illustrate how Practical Realism needs to be developed. He shows, on the one hand, that no version of the Standard View can accommodate collective beliefs, and, on the other hand, that Practical Realism can be applied to collective beliefs. But the application of Practical Realism to collective beliefs reveals a gap in the development of Practical Realism: What, exactly, *are* relational properties?

Before turning to Meijers’ major criticism—that I have not provided an adequate account of relational properties—I want to do two things: (1) Respond to a criticism; (2) Shed light on a point on which I think I was unclear in *Explaining Attitudes*.

(1) In applying Practical Realism to collective beliefs, Meijers says something with which I disagree. On my view, if S believes that p, then there are relevant true counterfactuals about what S would say, do or think in various circumstances. Meijers says rightly: “Given holism, where beliefs are part of a network of intentional states, the set of relevant counterfactuals is different for different agents have the same belief.”¹⁹ But then he goes on to say—and this is where I disagree—“This is another way of saying

¹⁹ This quotation and the others in this paragraph are from p. 10 of Meijers’ ms.

that every belief is always from a point of view.” I think that the matter of there being different sets of relevant counterfactuals for different agents having the same belief is not the same as the issue of point of view. Meijers objects that I say “virtually nothing in *Explaining Behavior* about the indispensability of a first-person point of view for the analysis of intentional states.” And he thinks that for Practical Realism, this is “an important issue since the specification of the conditions for having a particular belief are agent specific.”

By way of reply, I want to say two things. First, I do not think that the fact that the conditions for having a particular belief are agent specific implies anything about “the indispensability of a first-person point of view for the analysis of intentional states.” Agent specificity could be couched in wholly third-person terms: If S were in circumstances C, then S would do A. But, second, in fact, I do think that the idea of a first-person point of view is indispensable for understanding what a person is. Although, as Meijers notes, I did not discuss a first-person point of view in *Explaining Attitudes*, I am emphasizing the idea of a first-person perspective in subsequent writings.²⁰

(2) Meijers notes that I claim that whether an agent “has a particular belief is determined by what the agent does, says and thinks under various circumstances, where what the agent would do may be specified intentionally.”²¹ He calls this my “epistemological claim, since it is a claim about the attribution of beliefs.” Now I agree that such dispositions to say, think or act in certain ways are indicative of the global state of belief, but I think that such dispositions are more than indicative. On my view, such

²⁰ See “The First-Person Perspective: A Test for Naturalism” *American Philosophical Quarterly* 35 (1998): 327-348. An earlier version of this paper appeared in German in the proceedings of a 1997 conference at Humboldt University in Berlin on Naturalism. “Die Perspektive der Ersten Person: ein Test für den Naturalismus,” *Naturalismus. Philosophische Beiträge*, Geert Keil/Herbert Schnädelbach (Hg.), Frankfurt-am-Main (Suhrkamp Verlag). Finally, the idea of a first-person perspective is prominent in my account of persons in a forthcoming book, tentatively entitled, *What Am I? A Metaphysical Investigation into Persons and Bodies*.

²¹ This quotation and the others in this paragraph are from p. 10 of Meijers’ ms.

dispositions are metaphysically connected to belief. If one failed to have any dispositions to say, think or act in relevant ways, then one does not have the belief in question. Indeed, relevant dispositions are both necessary and sufficient (in a noncausal sense) for having a belief that p. Nevertheless, the fact that one is disposed to say or think certain things or to act in certain ways is not the same fact as the fact that one has a belief that p. There are other cases in which we say that A is necessary and sufficient (in a noncausal sense) for B; yet A's obtaining is not the same fact as B's obtaining. For example, the jury's voting 'guilty' is necessary and sufficient (in a noncausal sense) for the conviction of a defendant who pleads 'not guilty,' but the fact that the defendant is convicted is not the same fact as the fact that the jury votes 'guilty.' For one thing, the fact that the defendant is convicted is a fact about a particular person; and the fact that the jury voted 'guilty' is a fact about a group of people.

So, yes, my claim that whether an agent has a particular belief is determined by what the agent does, says and thinks under various circumstances is epistemological, but not *merely* epistemological. Although I admit that I could have been clearer on this point, I think that metaphysical and epistemological matters are connected in certain ways. It is not that I confuse metaphysics and epistemology; it is rather that I think that metaphysics cannot be done in a vacuum.

Now let me turn to Meijers' main criticism: that I do not provide an adequate account of relational properties.²² I confess that I do not. Indeed, I am not confident that anyone will come up with an adequate comprehensive account of relational properties. We tend to think of relational properties as expressible by two-place predicates like 'is taller than,' 'is the mother of,' and so on. But this tendency obscures a whole class of important properties that are relational in the sense that they could not be had by a being

²² On p. 12, Meijers suggests that I may hold that one can experience pain only if one stands in a certain relation to her environment. I would not go that far.

alone in the universe, but are not expressible by two-place predicates. The reason that they are not expressible by two-place predicates is that we have no idea how to specify exactly what relations must obtain for these relational properties to be instantiated.

Social and intentional properties are among such relational properties that are not expressible by two-place predicates. For example, the property of having tenure is perhaps a relation between a person and an institution. But can we say in nonintentional or physicalistic terms what an institution is, or what properties an institution must have for it to confer tenure? Or consider the property of being a registered voter. In order for someone to be a registered voter, there must be numerous complicated political and social conventions; but there is no way to specify what relations a person must have with a nonintentionally-described entity in order for the person to be a registered voter. Yet: we know that Smith has tenure, or that Jones is a registered voter, even though we have no idea of what relation Smith has to what, or what relation that Jones has to what, that makes it the case that Smith has tenure or that Jones is a registered voter. I am not sanguine that there is a single account of relations that can cover even these two cases, which are similar in that both involve intentional relations to institutions. I am even less sanguine that there will be a single account of relations that can cover all the kinds of relations that Meijers mentions: causal relations, intentional relations, normative relations.

I was hoping to start on the project of “giving relations their proper metaphysical due” by showing at least three kinds of things: (a) that many properties, beyond those expressible by two-place predicates, are relational in the sense that they could not be exemplified in a world in which only one individual existed; (b) that intentional states like belief depend on all kinds of relations, not specifiable in advance; and (c) that a Leibnizian account that reduces relational properties to intrinsic properties of relata is

unsatisfactory. A full account of relations was beyond the scope of my inquiry in *Explaining Attitudes*. However, investigation of relations per se seems to me an important research programme, and I would be happy if *Explaining Attitudes* stimulated such research.

(8) An Alternative to Both the Standard View and Practical Realism (Newen)

Professor Newen rejects (much of) the Standard View, and shares with me important philosophical views. For example, we both emphasize ways in which beliefs are context-dependent. Despite these similarities, however, Professor Newen proposes an alternative that he calls ‘Contextual Realism’ to both Practical Realism and the Standard View. Although he criticizes the Standard View for “disregarding the role of context in identifying the content of beliefs,” he also argues that “an adequate accommodation of the role of contexts (or contextual aspects) in describing the content of beliefs does not imply the denial of the main claim of the Standard view, that human beliefs are constituted by neural states.” (Newen, ms, 1)

First—and perhaps this is a minor point—Newen criticizes my analogy between beliefs/brain states and horserace-winning/leg states. Newen points out that ‘win’ is a success verb and ‘believe’ is not. Fair enough, but then in Newen’s own example of braking well, ‘braking well’ is a success verb too. Let me also register a related dissent about the first table in Newen’s paper. In the first table, Newen says that the (token) event of using the brake mechanism of the car during the period t is identical with the (token) event of decelerating the car during the period t . I would individuate events more finely, as Jaegwon Kim does, so that use of the brake mechanism would be a different event from deceleration of the car. Intuitively, one would think that use of the brake mechanism (as opposed to being struck by lightning) cause the deceleration of the car.

It is difficult to see to what extent Newen and I disagree because of two important terminological differences. We use the term ‘relational properties’ and the term ‘token’ in significantly different ways. Newen defines relational properties by “(rel)”, and says, “The intuitive content of this definition is that a property is relational only if it can be adequately characterized as a relation between two (or more) systems by using a part of the relevant predicate.” He then defines ‘context-dependent properties’ by “(con)” as follows: “The property *p* expressed with the monadic predicate ‘*p*’ is a context-dependent property of a system *S* if changing the context can modify the property expressed, i.e., in the context *c*₁ the property expressed by ‘*p*’ is *p* while in the context *c*₂ the property expressed by ‘*p*’ is the property *p*’.”

As I use the term ‘relational properties,’ the properties that Newen calls ‘context-dependent’ as well as the properties that he calls ‘relational’ are relational properties (in my sense). In *Explaining Attitudes* (p. 63), I define ‘relational properties’ like this: “*R* is a relational property if and only if: *x*’s having *R* entails that there is some *y* distinct from *x*.”²³ So, any property whose exemplification depends on a physical, social or linguistic context (context-dependent, in Newen’s sense) is, in my sense, a relational property. A property *P* is nonrelational if it supervenes on local microstructure in this sense: Necessarily, if *x* has *P* and *y* lacks *P*, then there is a microphysical difference between *x* and *y*. Any property (like having tenure, or believing that water is wet) that fails to supervene on local microstructure is relational in my sense. So when I say that beliefs are relational, I am saying something broader than what Newen takes me to be saying. If beliefs are context-dependent in Newen’s sense, then they are relational in my sense.

Under the heading “Points of Disagreement,” Newen says that I, like proponents of the Standard View, presuppose that “the natural basis of a belief could only be either a type of neural state or a particular neural token.” (Newen, ms, 15) He is correct that

²³ Also see my discussion of relational properties in my response to Meijers in the preceding section.

proponents of the Standard View suppose that if there are belief states, then they are identical to (or constituted by) types of neural states or particular neural states. But that is not my own view. When speaking from my own point of view, I do not refer to token beliefs at all. In fact I think that the type-token distinction as it is applied to beliefs in the literature is misguided in that it almost forces us to construe instances of beliefs as entities.

Perhaps I was not sufficiently clear, but one of the things that Newen says about Practical Realism I do not hold. Newen says: “The main point of Practical Realism is that beliefs cannot be characterized by internal states, because they are essentially relational.” (p. 3) Although it is my view that beliefs are not internal states; and it is my view that beliefs are essentially relational, I do not hold that the reason that beliefs are not internal states is that they are essentially relational. It is possible (although I think it false) that beliefs could be both relational and internal states. (That is what so-called “externalists” typically hold.) According to externalists, a belief is a brain state in an organism that has a certain relation to a certain environment. Believing that *p*, then, would be a relational property of a particular brain state. I do not connect being relational with not being an internal state in the way suggested by Newen’s sentence.

Newen’s view is that an instance of a belief is a particular neural state together with a context. He says that “the natural basis of a belief state (token) is neither a token of a special type of neural state alone nor a particular neural token alone but rather a complex token consisting of a neural state together with the context of the believer.” The context of the believer is what Newen calls “the internal and external context of a person, i.e., the individual learning history and the standard physical, social and linguistic environment of a person.”

Now Newen uses the term ‘token’ in an unusual way. A token (in the familiar sense) is a particular, something that occurs at a certain time in a certain place. (I eschew the term ‘token’ for beliefs, because I think that beliefs—unlike, say, baseballs—do not have precise locations.) But as Newen uses the term ‘token,’ a belief token does not seem to have a precise location either. Since the token is said to *consist*, in part, of the individual learning history, it does not occur at a single time or during a specifiable interval, but is temporally extended indefinitely into the past. (Who’s to say exactly when a person’s learning history (as opposed to a computer’s) began? Since the token is said to consist partly in the individual’s social and linguistic environment, it does not occur a single place, but is spatially extended indefinitely away from the believer. Except for the use of the word ‘token,’ I agree that these factors—neural state, learning history, social and linguistic environment—are relevant to what beliefs a person has. But if I am right, there will not be specifiably different neural states for different beliefs.

Newen illustrates his view with regard to neural nets. “The neural nets...may have the same cognitive state, i.e., recognizing a square, realized by neural states....The recognition is constituted by a certain net state together with the type of relevant context (the relevant context can be characterized by the connectivity, initial configuration and learning history of the net).” He continues: “Even if one belief is realized by different types of neural states it does not follow that beliefs are not constituted by neural states.” (Newen, ms, 15) Now I fully agree. My argument in ABBS is that it is unlikely that the requisite neural states to identify as beliefs will be found, and that if no neural states that are plausibly candidates to be beliefs are found, then the Standard View should be abandoned. But I am willing to wait and see.

Lynne Rudder Baker
University of Massachusetts/Amherst
January 26, 2001

