

Content and causation

Proponents of the Standard View who countenance beliefs – non-eliminativists – have the task of showing how beliefs may be scientifically respectable internal states suitable for causal explanation. Since beliefs are identified by content, the task is to show how content may be assigned to internal physical states in such a way that beliefs can be causally explanatory. Although assignment of content to brain states is a purely technical problem – and as a result, this chapter is a fairly technical discussion – it is a problem whose solution is required if there are beliefs as construed by the Standard View. Eliminative materialism, discussed in Chapter 3, is not faced with these problems since it does not recognize beliefs anyway; but the problem of content and causation is an urgent one for noneliminative proponents of the Standard View.

Without trying to survey all the recent work on content, I consider three different approaches to the problem of assigning content to internal states. Two appeal to a language-of-thought hypothesis. The first, proposed by William G. Lycan, tries to show that brain states are syntactically structured entities; the second, proposed by Jerry A. Fodor, looks to a new kind of semantic property – narrow content – to be causally explanatory. The third approach, Fred Dretske's, offers an account of belief as indication, without appeal to a language of thought. I argue that none of these theories is satisfactory: They all have technical (but interesting) difficulties that seem insoluble. All the views under consideration here share the Standard View assumption that beliefs must be physically realized internal states in order to have a causal-explanatory role in behavior. That assumption – the first premise in the Argument from Causal Explanation for the Standard View – is the target of Chapters 4 and 5.

Because I use thought experiments here, let me defend my method at the beginning. In this chapter, I am concerned with various theories of the contents of beliefs. Theories have modal, or counter-

factual, force: They purport to tell us not just what in fact has been the case or will be the case, but also what would be the case if certain conditions were to obtain. They tell us not only the conditions under which a person is in a state of a certain type, but they entail that if anyone *were* in those conditions, then she *would be* in a state of that type. In the thought experiments that follow, I imagine that two people are in the same conditions that the theory deems relevant, and I ask whether they are in the same state of the relevant type.

To isolate the features that the theories claim to be relevant, and to avoid extraneous issues, I hold constant features irrelevant to the theory – at the cost of empirical improbability. But the fact that it is empirically improbable, for example, that two people are ever in exactly the same (theory-relevant) conditions simply does not matter. What does matter is this: The thought experiments do not rip the concepts out of the contexts in which they have application; I am not supposing that it makes sense to ask, say, whether it is now five o'clock on the sun.¹ Rather, I am imagining the kinds of ordinary situations in which the concepts in question are designed to apply. Then, bracketing irrelevant features of the individuals in those situations, I ask: Do the dictates of the various theories accord with the core ways in which we actually apply the concepts in question? Since we do not have, for example, multiple subjects in the same brain states available for testing, we cannot actually control the conditions the theories take to be relevant to belief. In that case, it is difficult to see how better to test theories of belief than by imagining the theory-relevant conditions to be satisfied in ordinary situations, where there is no violation of known law or even abrogation of custom or convention. There is nothing unfamiliar about the situations envisaged by the thought experiments.

SYNTAX AND THE PROBLEM OF THE PARAMETER

Many philosophers hold that propositional objects of belief (expressed in 'that' clauses of attributions) have constituent structure. If someone believes that snow is white and believes that snow is

¹ Wittgenstein emphasized the senselessness of asking such questions.

² We may change application of a concept under pressure from a theory (e.g., we may begin to take nine-month-old fetuses to be persons on the basis of biological theory); but central cases are resistant to such pressures. Unless we were to give

cold, she has two beliefs with an element in common – a concept of snow. A number of philosophers of mind think that mental states themselves, not just their propositional objects, also have constituent structure. On the hypothesis of a language of thought, belief states have syntactic structure.³ Psychological processes are causal processes on sentence-like entities, individuated syntactically. The idea is that syntax is, as William G. Lycan says, “psychologically real,” that is, syntax is physically realized by structures in the brain. Because syntax is assumed to be fully determined by, or to be supervenient on, brain structure, syntactic properties are thought to be able to be causally explanatory.⁴

Since the language-of-thought hypothesis seems made to order to explain linguistic behavior, I focus on utterances as caused by brain states with syntactic structure. Let me make a simplifying assumption: Suppose that we identify certain utterances as ‘standard’ in the following way: They are sincere, assertive utterances that *p* that are causally explained by a belief that *p* and a desire to assert that *p*.⁵ Call the belief that *p* that causes a standard utterance that *p* the ‘related belief’. This assumption may be grossly implausible, but, as we shall see, even sticking with standard utterances, matters are not simple.

How is the syntactic structure of a belief to be determined? One prominent suggestion that is reasonably well worked out is Lycan’s: Associated with every sentence is a “semantic representation,” which displays the logical form of the sentence and which can serve as input for syntactic transformations. For each belief, say, that snow is white, there is a neural sentence that gives the logical form of the sentence ‘snow is white’. Now suppose that someone issues a standard utterance of ‘I am tired now’. What is the appropriate syntax to attribute to the related belief? One obvious suggestion is that the syntax has parameters for the indexical ele-

up applying the concept of a person altogether, we would continue to apply it to healthy adult human beings in the face of any new biological theory.

3 See, for example, Jerry A. Fodor, “Why There Still Has to Be a Language of Thought,” in *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge MA: MIT/Bradford, 1987), 135–54.

4 Note that this is an inference from “the order of being.”

5 It is noteworthy – although the literature does not suggest it – that it is highly unusual to cite a belief that *p* as causally explaining an assertion that *p*. Unless one doubts the speaker’s sincerity or linguistic competence, or doubts the truth of *p*, we just assume that one believes what one asserts.

ments. So, we must allow for indexical and other contextual elements to be provided for in the syntax of beliefs.

Call features of context that affect the truth conditions of a sentence or belief ‘semantically relevant features’. In public language, a hearer’s appreciation of context may contribute to the causal relevance of an utterance. For example, suppose that Jill’s shouting to Jack, “There’s a charging bull,” causes Jack to jump a fence. The utterance has its effect partly in virtue of its meaning and partly in virtue of the context in which it occurs: Jack understands Jill’s utterance; as an English speaker, he knows what it means, and being aware of the circumstances, he knows how to respond. With the language of thought, however, the situation is different. If Jack’s jumping the fence is caused by his belief that there’s a charging bull – a belief that he may have acquired by understanding Jill’s warning – the belief has its effect (of getting Jack out of the way of the bull) solely in virtue of the syntax of the internal sentence. (As Lycan says, concurring with Harman, “I do not ‘understand’ my own language of thought in the same sense in which I understand a natural language.”⁶) In the language of thought, those features of context that can make a difference in behavior must be explicitly represented in the brain. There is not a similar constraint on the representation of the logical forms of sentences in public languages, where salient features of context may be taken for granted.⁷

Since the causal efficacy of sentences in the language of thought, on this view, is wholly determined by brain states of the agent, every element whose presence or absence can affect behavior must be represented in the brain. So, whether or not public language requires that every semantically relevant feature be represented in logical form, the function of the syntax of the language of thought in causing behavior requires that every semantically relevant feature be explicitly represented in logical form and physically encoded in the brain.

So, the syntactic structure of any belief must have a “slot” (a parameter or a variable) for each semantically relevant feature. For example, in the case of “I am tired now,” the syntactic structure of the belief must have parameters for speaker and for time. Obviously, there are many other contextual elements that must be syn-

6 William G. Lycan, *Logical Form in Natural Language* (Cambridge, MA: MIT/Bradford, 1984), 237.

7 This was pointed out to me by Max Cresswell.

tactically represented in the brain state. For example, there are all sorts of hidden parameters in ordinary discourse ("I gave it to the woman on the left" – on the left of what?). The belief that standardly would cause assertion of "I gave it to the woman on the left" must have a "slot" for the hidden parameter to be filled in by whatever the woman was to the left of.⁸

For this picture to serve its purpose of showing how syntactically structured brain belief states cause behavior, it must impose two constraints on the syntactic structure of beliefs:

- (A) Syntactically distinct beliefs have physically distinct realizations in the brain.
- (B) A belief with n parameters is syntactically distinct from a belief with $n + 1$ parameters.

(A) is simply the requirement that syntax is to be physically represented in the brain. (B) may seem more controversial: After all, semantical indexes are not always marked in surface sentences; so why must they be marked in neural entities? Regardless of how we think of the semantics of public language, semantical indexes would still have to be marked in the language of thought if the language of thought is to play its causal role in producing behavior (as proponents of the Standard View construe causal role). The neural entity is supposed to be a syntactically structured inner cause: It must be able to cause behavior in virtue of its syntactic properties. Anything that can make a difference to behavior must be represented syntactically in the brain. Differences in number of parameters are differences in number of semantically relevant features; this kind of difference in truth condition can make a difference in behavior, and hence must be reflected in different physical realizations in the brain.

Unfortunately, (A) and (B) lead straight to trouble. The problem arises for any term that some people take to be relational, and others take to be nonrelational; terms of morals and manners come readily to mind. It also arises for terms, like 'tall', which apply to objects only relative to some implicit reference class; again, semantically relevant features that may remain implicit in a public language must be made explicit in the language of thought. I shall illustrate the general difficulty by an example from science since we

agree on its correct truth conditions, but I emphasize that the problem is not confined to theoretical contexts. Consider various standard utterances (all true and correctly believed to be true) of the following English sentence.⁹

- (s) An event on the sun is not simultaneous with anyone's seeing it.

The first utterance is by a nineteenth-century Newtonian physicist; the second utterance is by a twentieth-century Einsteinian physicist. Do the related beliefs, realized in the brains of the speakers, have the same syntactic structure or not?

(YES): Suppose so. If one assumes that Einstein's theory, with a frame-of-reference parameter for simultaneity, gives the actual truth conditions of (s), then the Einsteinian's related belief must represent all the semantically relevant features of (s) and the internal sentence must have a frame-of-reference parameter. So, if the Newtonian's belief has the same syntactic structure as the Einsteinian's, the Newtonian's related belief must also have a frame-of-reference parameter. Question: How did a frame-of-reference parameter get into the Newtonian's head?

In the unlikely event that we could answer that question, more are waiting in the wings. Since the Newtonian would deny that simultaneity was relative to inertial frame, must we conclude that the Newtonian did not understand Newtonian physics? What about Newton himself? Must we say that, in maintaining absolute simultaneity, he did not know what he was talking about? What are we saying when we say, "Newton believed that simultaneity was absolute"? And when Aristotle said, "Those things are called *simultaneous* without qualification and most strictly which come into being at the same time" (*Categories*, 14b25), did he have a frame of reference parameter in his head?

On the current alternative of saying that all the standard utterances of (s) have the same truth conditions, which are given by relativity theory, there is no way even to entertain the possibility of absolute simultaneity. For if all standard utterances of (s) have a frame-of-reference parameter, then it is difficult to see how even to

⁸ These ideas are taken from *Logical Form in Natural Language*.

⁹ This and other criticisms of Lycan's version of the language of thought hypothesis may be found in greater detail in my "Truth in Context," *Philosophical Perspectives* 2 (1989): 85–94.

formulate truth conditions for an assertion of absolute simultaneity.

If we assume that an expression used in a counterfactual utterance has the same meaning that it does in a standard utterance, the following becomes unintelligible (on the current alternative): "If Newton had been right, then simultaneity would have been absolute" (= 'it would have been the case that simultaneity is absolute').

Moreover, if the correct truth conditions for (s) are already encoded in Newton's brain, then the correct theories are already represented in our brains before they are 'discovered'. In that case, to find out about the physical world, we should not do physics, but psychology and linguistics. Thus, it seems hopeless to suppose that the beliefs related to standard utterances of (s) of the Newtonian and Einsteinian physicists are brain states with the same syntactic structure. So, turn to the other alternative.

(NO): Suppose that the physically realized beliefs related to the Newtonian's and Einsteinian's standard utterances of (s) have different syntactic structures. Then, presumably, the Einsteinian's brain state has a frame-of-reference parameter that the Newtonian's lacks. In that case, 'simultaneous' would be ambiguous. Of course, on some radical conceptions of theory change, it is ambiguous. Let us investigate the implications of ambiguity for the language-of-thought hypothesis.

To bring out the difficulty, I shall set up an example in a way that highlights exactly the relevant points. Suppose that the Newtonian and Einsteinian each has a daughter and a son; it happens that the daughters, who have been brought up in restricted environments, have been subjected to exactly the same kinds of sensory stimulation over their lifetimes, and that at the time we encounter them, their brains are in the same state. Suppose that both parents, emitting exactly the same noises, tell their respective daughters that it takes time for light to get to Earth from the sun, and hence that (s). Now, suppose that, in each scene, the brother arrives, and each daughter (flushed with new knowledge) issues a standard utterance of (s).

On the current alternative, brain states that are the beliefs related to Newtonian and Einsteinian physicists' respective standard utterances of (s) differ in truth condition: The Einsteinian's, but not the Newtonian's, brain state has a frame-of-reference parameter. If we also assume that their daughters' similar beliefs inherit their parents' truth conditions (otherwise, it would be miraculous that

anyone ever learned a language), then at least one of the constraints (A) or (B), is violated. Since the daughters' brains were in the same states and they had the same physiological histories, and the information about simultaneity was transmitted by means of physically identical sounds, they do not now have different physical realizations in their brains; so by (A), their brain states are syntactically the same. But since, by assumption (NO), one brain state has a frame-of-reference parameter but the other doesn't, their beliefs have syntactic structures with different numbers of parameters; hence, by (B), their brain states are syntactically distinct. Hence, (A) and (B) lead to contradiction.

Lycan has replied to this by holding, at least tentatively, that "simultaneous" remains a two-place predicate for ordinary people but has become a three-place predicate for the *cognoscenti*.¹⁰ Then, since neither daughter knows physics, "presumably neither has the three-place predicate on board." So, according to Lycan's tentative reply, the beliefs that produce standard utterances of (s) in the two girls do have the same syntactic structure. On the other hand, according to Lycan's theory, syntactic structures realized in the brain encode the truth conditions of the belief. If, however, 'simultaneous' is a two-place predicate for both daughters, then the girls' standard utterances of (s) do not have the correct truth conditions of (s): For, as we now know, simultaneity is relative to frame, and hence 'simultaneous' is in fact a three-place predicate. Thus, Lycan's tentative reply seems to entail that the girls' standard utterances of (s) have truth conditions that in fact they do not have.

Moreover, suppose that, as Lycan's reply suggests, the surface sentence, 'An event on the sun is not simultaneous with anyone's seeing it', is ambiguous: Underlying utterances of it are two logical forms with two different truth conditions. It is difficult to see how to understand the ambiguity. Contrast a typical case of ambiguity – for example, 'I put my money in the bank'. The ambiguity in such typical cases arises from the fact that an occurrence of a surface expression like $b^{\wedge}a^{\wedge}n^{\wedge}k$ may really be either of two different words: 'Bank' may refer to a financial institution or to the ground beside some river or stream. But 'simultaneity' is not like 'bank': There is only one physical phenomenon of simultaneity, and it is relative to frame. So, if we assume that 'p' is true if and only if p is the case,

there is only one set of truth conditions for utterances of (s), and in it 'simultaneous' is a three-place predicate. 'Simultaneous' thus should not be taken as ambiguous. If each daughter represents 'simultaneous' by a two-place predicate, then neither daughter represents the truth conditions of (s) – in violation of Lycan's theory.¹¹

Finally, any strategy that takes different people's standard utterances of (s) to have different truth conditions inherits all the difficulties attending the thesis in the philosophy of science that theories are incommensurable – and then compounds them by supposing that the incommensurable relations between theories must be reflected in the physical structures of the brain. On the current alternative, the Newtonian says nothing false when she says, "Simultaneity is not relative to a frame of reference." For the Newtonian's utterance has different truth conditions from the Einsteinian's; on the current alternative, the Newtonian's utterance is true if and only if simultaneity construed as a two-place relation is not relative to frame; so, the Newtonian's utterance is true. Moreover, on the assumption that their utterances of (s) have different truth conditions, the Einsteinian and the Newtonian don't disagree when one says, "Simultaneity is absolute," and the other says, "Simultaneity is relative." Lycan's reply seems to subject his view to all the counterintuitive consequences of the "incommensurability" view of theory change.

Therefore, it seems that either answer to the question – Do standard utterances of (s) by both Newtonians and Einsteinians have the same truth conditions? – comes to grief when we assume that syntax is "psychologically real." So, I think that the language-of-thought hypothesis is afflicted with what we might call 'the problem of the parameter.'¹²

Actually, to use an example like 'simultaneity' underestimates the problem of the parameter for the language-of-thought hypothesis. For, as already suggested, the problem of the parameter may be generated by any hidden parameter: Consider (putative) representa-

11 Perhaps one would want to relativize truth conditions to individuals: Conditions under which (s) is true for one may not be the same as the conditions under which (s) is true for another. Not only is this an implausible construal of truth conditions of a public language, but also it is not available to realists like Lycan. 12 The example has implications beyond the language-of-thought hypothesis. First, it suggests that there are syntactic as well as semantic issues of theory change; second, it suggests that any theory of meaning based on syntactic primitives is at least liable to the problem of the parameter.

tions of 'slurping soup is impolite' in the heads of an absolutist and a relativist. In the scientific case, theories provide accounts of which features are the semantically relevant ones. But in most ordinary contexts, things are not so tidy. We have no general theory of semantically relevant features of standard utterances, nor will we until we solve the frame problem.¹³

And until we do solve the frame problem, it seems to me that we have no theory whatsoever, only a relabeling of the problem. To say, as Lycan does, that all contextual elements are handled via an all-purpose assignment function is not to say anything informative without some account of what features are semantically relevant in general. Without such an account, the only available specification of the assignment function is that it takes as arguments all the different parameters realized in the brain and returns as values all and only semantically relevant features of the context, whatever they may be.

Note that I am not talking here about empirical questions of how the brain functions to encode the requisite parameters, nor am I asking for mechanisms that show how the brain computes the values for the assignment function. I am asking the prior question of what parameters need to be encoded (by whatever mechanisms). For the language-of-thought proposal to work, we need a context-free theory of context, but I see no such theory in the offing.¹⁴ Lycan remains unmoved: "Our present or even future inability to specify a context does not matter to the thesis that a determinate assignment function exists."¹⁵ Perhaps; but in view of the difficulties presented here, I see no good reason to think that there is a function (in the mathematical sense) from particular brain states to the semantically relevant features of utterances.

In any case, the problem of the parameter would seem to afflict any account of beliefs as syntactically structured brain states. Since any language-of-thought hypothesis assumes that syntax is physi-

13 The frame problem is how to get a machine to update knowledge of a changing situation by "noticing" salient features and ignoring others. For example, knocking the support out from under a box causes it to change position but not to change color. See Hubert L. Dreyfus, *What Computers Can't Do*, rev. ed. (New York: Harper and Row, 1979) and John Haugeland, *Artificial Intelligence: The Very Idea* (Cambridge, MA: MIT/Bradford, 1985).

14 Since I do not believe that we will ever have a context-free theory of context, I am prepared to live with a measure of context dependence – in the evaluation of counterfactuals, for example.

15 Lycan, "Reply to Baker," 99.

ally realized in brains and any such hypothesis must accommodate semantically relevant features of context in some way, the problem of the parameter is not peculiar to Lycan's view. Rather, any view that takes the syntax of internal sentences to cause behavior must have some way to specify the syntactical features of the language of thought that avoids these difficulties. These examples show that we can not simply assume that it makes sense to think of the brain as organized in terms of states that generally realize syntactic properties.

THE DEAD END OF NARROW CONTENT

Beliefs are ordinarily attributed in English by sentences with embedded 'that' clauses – for example, 'Jones believes that rock-climbing is dangerous'. Such sentences identify beliefs by what have come to be called 'broad contents'. Since broad contents individuate beliefs in part by reference to the believer's environment, beliefs are *relational* mental states: The conditions for having a belief, say, that water is wet or that arthritis is painful, depend not only on the intrinsic properties of the believer, but also on the nature of the believer's physical and social environment.¹⁶

Assuming that beliefs individuated by 'that' clauses (or by broad contents) are relational, I am here concerned with the causal-explanatory status of belief states. Are beliefs (or the properties that individuate them) causally explanatory?¹⁷ Are relational properties ever causally explanatory? Some philosophers – prominently, Jerry A. Fodor – acknowledge the causal relevance of relational properties generally, but take beliefs individuated by broad content to be metaphysically unsuitable for purposes of causal explanation.¹⁸ I

16 Tyler Burge is largely responsible for the widespread agreement that (*de dicto*) beliefs as ordinarily attributed are relational. See his "Individualism and the Mental," in *Studies in Metaphysics* (Midwest Studies in Philosophy, 4), ed. Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein (Minneapolis: University of Minnesota, 1979), 73–122.

17 I follow Fodor and speak sometimes of states and sometimes of their individuating properties as causally explanatory.

18 See *Psychosemantics* and "A Modal Argument for Narrow Content," *Journal of Philosophy* 88 (1991): 5–26. Hereafter, references to this article appear in the text as "MANC," followed by a page number. In "The Elm and the Expert: Mental-essence and Its Semantics" (1993 Jean Nicod Lectures), Fodor gives up his theory of narrow content. My arguments here remain important, however, in that they show that denial of the causal explanatory status of broad content leads to denial of

want to challenge this position, by arguing that broad contents are causally explanatory, and that if Fodor's argument were to cast doubt on the claim that broad contents are causally explanatory, then it would cast doubt equally on the claim that any relational property is causally explanatory.

Explanatory properties, according to Fodor, are taxonomic, that is, they are projected by the laws of some science; and since the sciences aim at causal explanations, Fodor holds, taxonomy in the sciences is by causal powers.¹⁹ Fodor argues that broad contents do not contribute in the relevant way to an individual's causal powers, and hence that they can not be taxonomic in psychology. Nonetheless, he upholds the explanatory status of other relational properties; indeed, Fodor says, "Taxonomy by relational properties is ubiquitous in the sciences" (MANC, 12). Thus, Fodor defends the conjunction of (A) and (B):

(A) Relational properties that individuate belief states are not taxonomic in psychology.

(B) Some relational properties are taxonomic in the special sciences.

I try to show here that (A) and (B) do not sit comfortably on the same bench. Fodor's arguments, I urge, either fail to disqualify broad contents as taxonomic, or else disqualify all relational properties as taxonomic. I am not going to claim that broad contents must be taxonomic in psychology, only that the metaphysical considerations against their being taxonomic are faulty.²⁰ Logically and metaphysically speaking, as broad contents go, so go relational properties generally – Fodor's claims to the contrary notwithstanding.

the causal explanatory status of relational properties generally. In Chapter 6, I criticize Fodor's latest proposal.

19 Sometimes Fodor speaks of causal powers as properties (as in "a cause property might fail to count as a causal power in virtue of its responsibility for one effect property, but still might constitute a causal power in virtue of its responsibility for some other effect property"); and sometimes he speaks of causal powers as the things that have such properties (as in "We have seen that water thoughts and water thoughts are not different causal powers"). For my purposes, I think that I can overlook this ambiguity. MANC, 12, 25.

20 For a different argument, see Robert van Gulik, "Metaphysical Arguments for Internalism and Why They Don't Work" in *Representation: Readings in the Philosophy of Mental Representation*, ed. Stuart Silver (Dordrecht, Holland: Kluwer Academic Publishers, 1989), 151–9.

Fodor's argument for (A) in MANC is part of an argument that intentional psychology individuates states with respect to narrow content, where narrow contents are nonrelational. Narrow content supervenes on the subject's intrinsic properties, without regard to the subject's environment. The skeleton of Fodor's new argument for narrow content is this:

- (1) All scientific taxonomies individuate states with respect to their causal powers.
- (2) Intentional psychology individuates states with respect to intentional content.
- (3) Difference in broad content does not suffice for (relevant) difference in causal powers.

Therefore,

- (4) Intentional psychology individuates states with respect to narrow content.

Fodor's latest argument consists mainly of a new defense of (3), in which Fodor proposes a necessary condition (what I shall call the "no-conceptual-connection" test) for a difference to count as a difference in causal power, and then claims that broad contents fail it.

Fodor formulates two tests – the no-conceptual-connection test and the cross-context test – for determining when a property is a causal power and hence may be taxonomic in some science. More precisely, the tests are to show when the difference between having a particular property and not having it is a difference in causal power, in virtue of the responsibility of the property for properties of the subject's behavior. Since Fodor holds that taxonomic properties in psychology must make a difference to the subject's actual or possible behavior, only properties whose possession makes a difference to the bearer's causal powers can be taxonomic. Fodor argues that broad contents can not be taxonomic in psychology, because they fail the no-conceptual-connection test, but that other relational properties, like the property of being a planet, can be taxonomic in other sciences, because they pass both tests. I argue that the only principled way that Fodor has to rule out broad contents as taxonomic would also rule out other relational properties, like that of being a planet, as taxonomic. In particular, broad contents actually pass Fodor's no-conceptual-connection test in the relevant way; and

any interpretation of the cross-context test which would disqualify belief individuated by broad content as taxonomic would also disqualify relational properties generally as taxonomic.

The no-conceptual-connection test

I want to show that broad contents in fact do satisfy Fodor's necessary condition for a difference to count as a difference in causal power, in virtue of its responsibility for a difference in behavior.²¹ Hence, the argument for (3) collapses.

Fodor offers a schema in terms of which he casts his argument. Consider a situation in which there is a pair of causes C_1 , C_2 , and their effects E_1 , E_2 , such that

C_1 differs from C_2 in that C_1 has cause property CP_1 , where C_2 has cause property CP_2 ,
 E_1 differs from E_2 in that E_1 has effect property EP_1 and E_2 has effect property EP_2 .

The difference between C_1 and C_2 is responsible for the difference between E_1 and E_2 in the sense that, if C_1 had had CP_2 rather than CP_1 , then E_1 would have EP_2 rather than EP_1 ; and if C_2 had had CP_1 rather than CP_2 , E_2 would have had EP_1 rather than EP_2 . (MANC, 9)

I shall follow Fodor and think of the schema "sometimes as relating events and sometimes as relating event types" (MANC, 9). Now, asks Fodor, which instances of the schema "are cases where the difference between having CP_1 and having CP_2 is a difference in causal power in virtue of its responsibility for the difference between E_1 and E_2 ?" More briefly, when is the difference between CP_1 and CP_2 a difference in causal power? Fodor's answer: when the difference in cause properties is not conceptually connected to the difference in effect properties.²² Fodor initially states this requirement by saying that the difference between CP_1 and CP_2 is a difference in causal power

only when it is not a conceptual truth that causes which differ in that one has CP_1 where the other has CP_2 have effects that differ in that one has EP_1 where the other has EP_2 . (MANC, 19)

²¹ Fodor is specifically concerned with properties' being causal powers in virtue of their responsibility for the properties of their bearers' behavior, but I shall leave this qualification implicit in most of what follows.

²² I do not explore complexities (and perplexities) surrounding the notion of conceptual connection here. The argument to follow applies to any account that takes conceptual truth to be a species of necessary truth.

This necessary condition is supposed to rule out broad contents as causal powers because, although the difference between having water thoughts and having twin-water thoughts is responsible for the difference in intentional properties of behavior (e.g., drilling for water versus drilling for twin water), it is a conceptual truth that thoughts that differ only in being water or twin-water thoughts have effects that differ only in being water drillings or being twin-water drillings.

The necessary condition is then revised to take care of an objection posed by Stephen Stich. Suppose, for example, that water is Bush's favorite drink. Then, it is not a conceptual truth that beliefs that differ in that one is about Bush's favorite drink and the other is about twin water have effects that differ in that one is a water behavior and the other a twin-water behavior. Yet, Fodor does not want to count the difference between being about Bush's favorite drink and being about twin water as a difference in causal powers. So, he must amend his necessary condition on causal powers to rule out such a case.

Although Fodor never actually formulates the patched-up version of the necessary condition, he adds the requirement that the following not be conceptually necessary:

If *B* [e.g., being concerned with Bush's favorite drink] is a property that water behaviors have, then if my thoughts are water thoughts, then my behaviors have *B*.

If *B* is the property of being concerned with Bush's favorite drink, then this instance of the conditional is, as Fodor wants, conceptually necessary. There is a conceptual connection between water thoughts and water behaviors, and there is no possible world in which being concerned with Bush's favorite drink is a property of water behaviors, and Fodor's thoughts are water thoughts, yet Fodor's behaviors fail to be concerned with Bush's favorite drink.

To see that not all conditionals of this form are conceptually necessary truths, Fodor says, suppose that thinking about topology causes headaches and compare the following conditional:

If *B* [e.g., being painful] is a property of headaches, then if *S*'s thought is about topology, *S*'s mental state is painful.

This conditional, though true, is not a conceptually necessary truth. The "headache" conditional is only a contingent truth, because it is a contingent truth (if it is a truth at all) that topology

thoughts cause headaches. In some other world, headaches have *B* (the property of being painful), *S* has topology thoughts, yet *S*'s mental state lacks *B* — because in that world topology thoughts do not cause headaches.²³

Putting these conditions together, we have the following as a necessary condition on causal powers: Suppose that C_1 , C_2 , CP_1 , CP_2 , E_1 , E_2 , EP_1 , and EP_2 satisfy Fodor's schema. Then:

(FCP) Two cause-properties, CP_1 and CP_2 , are different causal powers only if neither (i) nor (ii) is a conceptual truth:

(i) Causes C_1 and C_2 , which differ in that C_1 has CP_1 and C_2 has CP_2 , have effects, E_1 and E_2 , which differ in that E_1 has EP_1 and E_2 has EP_2 ; and

(ii) If *B* is a property that events with EP_1 have, then if C_1 has CP_1 , then E_1 has *B*.²⁴

Fodor believes that broad contents fail to satisfy this necessary condition for being causal powers. I believe, however, that Fodor has too narrow a view of the difference that difference in broad content can make. I want to show that differences in broad content unaccompanied by physiological differences make a causal difference in behavior that satisfies Fodor's necessary condition(s) on causal powers. So, let us consider an example.

In English, the word 'jade' denotes both jadeite and nephrite, which differ in structure. Although they are similar in appearance, jadeite, which is found mainly in Burma and in Central America, is much more valuable than nephrite, which is found all over the world. Despite the fact that knowledgeable people are aware of these differences, 'jade' in English still refers not only to jadeite but

23 I am paraphrasing Fodor here. I believe that this point raises deep questions about (Fodor's conception) of the nature of causal laws. In "Making Mind Matter More," *Philosophical Topics* 17 (1989): 63, Fodor says that he is "hard put to see how anybody could seriously object" to the "idea that hedged (including intentional) laws necessitate their consequents when their ceteris paribus clauses are discharged." Fodor's conception of causal law there warrants close attention, which I can not give here.

24 Since Fodor himself never actually formulates his amended condition, I can only guess at how clause (ii) should go. Clause (ii) as stated needs further work; however, since I do not see how better formulation of (ii) could block my counterexample, I do not undertake to improve it here. Ultimately, it is up to the proponent of the "no-conceptual-connection" test to formulate the condition that is supposed to block counterexamples.

to nephrite as well. As (spoof) proof, let me cite *Webster's Unabridged Dictionary*, which gives the following as a definition of 'nephrite': "the less valuable of two varieties of jade, compact in structure and varying in color from white to dark green."

Now suppose that there is another community, in which all the differences between jadeite and nephrite are also well known by the experts and by the informed jewelry-buying public. But in the other community, the word that sounds like 'jade' denotes only jadeite. 'Jade' is as inapplicable to nephrite in the other community as 'gold' is to iron pyrite in our community. The less valuable nephrite is called something else and is not in the extension of 'jade'. The truth conditions of the sentences 'There are jade stones', then differ in the two communities. In English, 'There are jade stones' is true if there are either jadeite or nephrite stones. In the other community, 'There are jade stones' is true if there are jadeite stones.

Consider another possible world that has in it both our English-speaking community and the other community (or their counterparts), and suppose that there are two microphysical duplicates, Ann and Jan, in that other world. Ann lives in the English-speaking community, and Jan in the other community. Although both use (what sounds like) 'jade' in various correct sentences in their respective communities, neither is a jewelry buyer, and neither knows that there are two similar kinds of tough green stones. So, when Ann has thoughts about jadeite or nephrite, Jan has thoughts about jadeite.

Now suppose that Ann and Jan both appear as contestants, in their respective communities, on qualitatively identical quiz shows. For the grand prize, each has to identify a stone. (The stones are qualitatively identical pieces of nephrite.) Each quiz show host says: "Here's a lovely green stone. Can you identify it?" To this, Ann and Jan give acoustically identical replies: "The stone is jade." Now Ann has given a winning answer, in Ann's community, nephrite is a variety of what is called 'jade'. Jan, however, has given a losing answer; in Jan's community, nephrite is not a variety of what is called 'jade'. At this point — when Ann hears the audience applaud and Jan hears the audience groan — the physical descriptions of the contestants part ways; Ann and Jan cease to be duplicates.

Now put this story into Fodor's schema. Ann tokens neuro-physiological type *T*, which has the property of being a belief with

the same truth condition as the English sentence, "The stone is jadeite or nephrite," and which causes emission of a sound of acoustical type *U* that has the property of being a winning answer. Jan tokens neurophysiological type *T*, which has the property of being a belief with the same truth conditions as the English sentence, "The stone is jadeite," and which causes emission of a sound of acoustical type *U* that has the property of being a losing answer. Schematically:

- (C₁) a state realized by neurophysiological type *T*
- (C₂) a state realized by neurophysiological type *T*
- (CP₁) being a belief with the same truth condition as the English sentence "The stone is jadeite or nephrite."
- (CP₂) being a belief with the same truth condition as the English sentence "The stone is jadeite."
- (E₁) emission of a sound of acoustical type *U*
- (E₂) emission of a sound of acoustical type *U*
- (EP₁) being a winning answer
- (EP₂) being a losing answer

The beliefs, (C₁) and (C₂), have different cause properties, (CP₁) and (CP₂), and the answers, (E₁) and (E₂), have different effect properties, (EP₁) and (EP₂). Furthermore, if (C₁) had had the truth conditions that (C₂) had, then (E₁) would have been a losing (rather than a winning) answer. Fodor proposes to block broad contents as causal powers if the relevant instances of (i) and (ii) in (FCP) are conceptual truths. To see that they are not conceptual truths, let (i') and (ii') illustrate relevant instances of (i) and (ii):

- (i') Two states realized by neurophysiological states of type *T*, which differ in truth conditions (as described), have effects (acoustically identical sounds), which differ in that one is a winning answer and the other is a losing answer.
- (ii') If being a winning answer is a property of winning answers, then if a state realized by neurophysiological state of type *T* has the property of being a belief with the same truth conditions as the English sentence, "The stone is

jadeite or nephrite', then the emitting of a sound of acoustic type *U* is a winning answer.

(i') and (ii') are, of course, true; but, obviously, they are not conceptually necessary truths. There is no conceptual connection between certain truth conditions and being a winning answer. (ii') is parallel to Fodor's "headache" conditional; the "headache" conditional is only contingently true, because in other worlds, topology thoughts do not cause headaches, and the conditional is false in such worlds. (ii') is only contingently true, because in other worlds, a belief with the given truth conditions may not produce a winning answer. If Ann and Jan had been presented with the same stones and asked the same questions in some context other than a quiz show, then their answers would not have had the properties, respectively, of winning and losing.

Although in the example, being a winning answer and being a losing answer are different "effect properties," notice that they, in turn, are causally efficacious and that they produce very different results. (E_1) elicits cheers from the studio audience; (E_2) elicits groans from the studio audience. (E_1) leads to Ann's taking away the grand prize; (E_2) leads to Jan's going away empty-handed. When Ann's husband suspiciously questions Ann about how she suddenly acquired such wealth, Ann can cite this as a cause: "I gave the winning answer." The winning answer allows Ann to retire while the losing answer forces Jan to return to a dreary job. Such differences in subsequent effects indicate that the quiz show episodes are parts of causal processes.

Someone may object that psychologists are not concerned with the difference between being a winning answer and being a losing answer. To this objection, I have a twofold reply. First, there may well be contexts in which the difference between winning and losing answers is exactly what a psychologist is interested in. (You may seek out such a psychologist to treat your underachieving child, who oddly produces wrong answers when you suspect that he knows better.) We can not say a priori under what kinds of descriptions psychologists will explain behavior. Second, even if psychology never countenanced properties like 'being a winning answer', that fact would be irrelevant to my point. For Fodor's condition is perfectly general; it does not apply exclusively to psy-

chology. My point is that Fodor gave a necessary condition for properties to be explanatory, and he claimed that broad contents do not satisfy this condition (the no-conceptual-connection test). And I provided a counterexample to show that, in fact, broad contents do satisfy it. Hence, Fodor's no-conceptual-connection test can not rule out differences in broad contents as causally explanatory differences, regardless of what counts as behavior in psychology.

Here, then, is what I claim for the "jadeite" example: Differences in broad content, unaccompanied by neurophysiological differences, causally explain differences in behavior that are not conceptually connected to the broad contents that explain them. Fodor may insist that such differences in broad content as I have described fail to be differences in causal powers in his sense; in that case, I would reply that causally explanatory properties need not be causal powers in his sense. (I do not care about the term 'causal powers', which seems to flop around anyway.) On the other hand, if we simply agree to call causally explanatory properties "causal powers," I do not believe that Fodor's conditions have ruled out broad contents as causal powers. Thus, I do not believe that Fodor has given reason to think that differences in broad content, unaccompanied by neurophysiological differences, fail to be causally explanatory. Fodor's necessary condition on causal powers, in terms of conceptual connections between cause and effect properties, does not preclude differences in broad contents as differences in causal powers. I now want to show that broad contents are as worthy as nonpsychological relational properties to be causally explanatory.

The cross-context test

The no-conceptual-connection test is not the only weapon against broad content in Fodor's arsenal. In addition to passing that test, explanatory properties must also pass the cross-context test. To see whether causal powers are the same or different, we must compare the individuals "across contexts rather than within contexts" (MANG, 8). The idea of the cross-context test is that two individuals have the same causal powers if and only if, in the same context, they have the same effects. To see whether a relational property makes a difference to causal powers, consider two individuals who are similar except that one has the relational property in question

and the other lacks it. Now, according to Fodor, the property makes a difference to causal powers only if the individuals have differed effects when considered "across contexts."

Before examining the cross-context test, note that Fodor himself no longer puts stock in the cross-context test to rule out broad contents as explanatory. Indeed, in order to motivate the non-conceptual-connection test, Fodor says that broad contents *do* survive the cross-context test:

Whatever the context of utterance, my utterance is a water request and his utterance is a twater request. So our behaviors remain relevantly different under these intentional descriptions *even by the across-context test*. It is this residual difference between the behaviors – their cross-context difference under certain intentional descriptions – which is the challenge to individualism and local supervenience. (MANC, 8–9; emphasis in original)

Fodor seems to be admitting here that – without the aid of the non-conceptual-connection test, which we have seen to be no help – the cross-context test does not disqualify differences in truth condition alone as differences in causal powers, in virtue of the effects of such differences on the properties of behavior. Nonetheless, the cross-context test deserves consideration. In particular, does my counter-example pass the cross-context test?

The difficulty with the cross-context test is that Fodor never explicitly formulates it, and his comments about it suggest more than one interpretation. I offer several interpretations of the cross-context test – all the interpretations for which I find evidence in Fodor – and argue with respect to each one of them either that broad contents pass it or that other relational properties like being a planet fail it (or both). If that is right, then the cross-context test can not rule out broad contents as taxonomic without also ruling out nonpsychological relational properties in good standing.

Here is an initially plausible way to interpret the cross-context test:

(CCT₁) Property *P* is causally explanatory only if its possession makes a difference to the causal powers of its possessor, where *x* and *y* have the same causal powers if and only if: if *x* had been substituted for *y*, then *x* (in *y*'s context) would have had all the same effects that *y* did have.

(CCT₁) would disqualify all relational properties – such as being a planet – as causal powers. Fodor says that being a planet is a rela-

tional property in good standing, and that this property could "distinguish molecularly identical chunks of rock," and that being a planet constitutes a causal power in good standing (MANC, 12). Let *R*₁ be a planet revolving around a star and *R*₂ be a nonplanetary microphysical duplicate held (for a time at least) in an elliptical orbit by the distribution of matter in the universe. Now substitute *R*₁ for *R*₂ and vice versa. Since *R*₂ is a microphysical duplicate of *R*₁, when *R*₂ is substituted for *R*₁, *R*₂ will orbit around *R*₁'s star, and hence will be a planet when put in *R*₁'s environment. So, if *R*₂ is substituted for *R*₁, *R*₂'s effects have all the same properties that *R*₁'s effects did have. Hence, on (CCT₁), *R*₁ and *R*₂ have the same causal powers, and the difference between being a planet and not being a planet fails to be a difference in causal power. So, on (CCT₁), the property of being a planet does not pass the cross-context test.

No relational property can pass the cross-context test as interpreted via (CCT₁) for the simple reason that (CCT₁) amounts to a requirement (or stipulation) that properties that suffice for a difference in causal powers be nonrelational. Indeed, (CCT₁) is almost a paraphrase of Stephen Stich's replacement argument for his "Autonomy Principle," the point of which is to confine explanatory properties to those that supervene on the current intrinsic properties of their bearers.²⁵ And, as Fodor points out, the property of being a planet does not supervene on the current intrinsic properties of its bearer. Thus, if we use (CCT₁) to interpret the cross-context test, then no relational properties are taxonomic. Since Fodor says that "taxonomy by relational properties is ubiquitous in the sciences," (CCT₁) does not yield the correct interpretation of the cross-context test.

So, let us try another interpretation. In discussing the cross-context test in "A Modal Argument for Narrow Content," Fodor comments in a footnote: "One applies the cross-context test by asking whether Ann would have the same effects as Jan does have if Ann were to interact with the same things . . . with which Jan does interact" (MANC, 8). This suggests interpreting the cross-context test by means of (CCT₂):

(CCT₂) Property *P* is causally explanatory only if its possession makes a difference to the causal powers of its possessor,

²⁵ Stephen P. Stich, *From Folk Psychology to Cognitive Science: The Case against Belief* (Cambridge MA: MIT/Bradford, 1983).

where x and y have the same causal powers if and only if: if x had interacted with the same things that y did in fact interact with, then x would have had all the same effects that y in fact did have.

If we interpret the cross-context test on the basis of (CCT₂), a difference in truth conditions suffices for a difference in causal powers, but the difference between being a planet and not being a planet does not suffice for being a difference in causal powers.

The case of broad contents: Ann and Jan also differ in causal powers if the original story is amended slightly. Suppose that Ann and Jan had never interacted with either jadeite or nephrite, and that both learned what sounds like 'jade' in their respective languages from teachers who had never interacted with either jadeite or nephrite either. Indeed, the teachers themselves could be microphysical duplicates. (If this seems implausible, take the original story and suppose that Ann and Jan have microphysically identical children to whom Ann and Jan teach what each calls 'jade' in her respective language – before the quiz show, while Ann and Jan are still duplicates.) The physical identities of the individuals with whom Ann and Jan interacted are irrelevant to the intentional and semantic properties that Ann and Jan acquire. So, given (CCT₂), the cross-context test does not block the counterexample.

The case of the property of being a planet: Again, let R_1 be a planet revolving around a star and R_2 be a nonplanetary microphysical duplicate held in an elliptical orbit by the distribution of matter in the universe. Then, if we assume gravitational pull to be an interaction, if R_2 had interacted with everything that R_1 in fact interacted with, R_2 would be revolving around the star and thus would be a planet; and if R_1 had interacted with everything that R_2 in fact interacted with, R_1 would not be revolving around a star and hence would not be a planet. Hence, on (CCT₂), the difference between being a planet and not being a planet fails to be a difference in causal powers.

So, (CCT₂) can not provide the interpretation of the cross-context test that suits Fodor's purposes: For (CCT₂) both allows the counterexample to go through and disqualifies differences in nonintentional relational properties like that of being a planet as differences in causal powers.

Here is a final attempt to interpret the cross-context test. In

introducing the cross-context test, Fodor gives an example in *Psychosemantics*: "Roughly, our biceps have the *same* causal powers if the following is true: For any thing x and any context C , if you can lift x in C , then so can I, and if I can lift x in C , then so can you."²⁶ This suggests interpreting the cross-context test by means of (CCT₃):

(CCT₃) Property P is causally explanatory only if its possession makes a difference to the causal powers of its possessor, where x and y have the same causal powers if and only if: There is no context C such that x has an effect in C that y in C does not have.

If we interpret the cross-context test on the basis of (CCT₃), Ann and Jan clearly have different causal powers. Here is a relevant context: Let the quiz show be part of the International Quiz Show Olympics, in which multilingual translators determine what each contestant says – what answer she gives. The contestants enter identical isolation booths and simultaneously have identical auditory sensations, but because of the differences in languages, they are not asked the same question. (Perhaps instead of a human translator, there is only a machine translator. The program of the machine includes, for each contestant, specification of the language that she speaks; so, each contestant's vocal emissions are automatically treated as being in her native language.) Physically speaking, there is a single context: a room with a translation device and two isolation booths and appropriately placed mirrors, so that the visual sensations of the contestants are also alike. In this context, the property of being a winning answer is one that Ann's behavior has and Jan's behavior lacks. Assuming that the quiz show's translators are competent and alert, Ann's is the winning answer, and Jan's the losing answer – as in the original story. So, given (CCT₃), the cross-context test does not block the counterexample.

Although I think that this is an adequate response to (CCT₃), let me elaborate a bit by posing a possible objection. The objection is that we should allow Ann and Jan to differ only in the truth conditions of their mental states, not in any other way that quiz show judges can detect.²⁷ This objection amounts to an ad hoc stipula-

²⁶ Fodor, *Psychosemantics*, 35.

²⁷ Paul Boghossian made a similar objection in conversation. I formulated the response that I give later in part in correspondence with Pierre Jacob.

tion. Typically, in Olympic competitions, the knowledge of national identities of the participants is highlighted, not bracketed. In any case, a counterexample may assume whatever is necessary for there to be a difference in truth conditions in what sounds like 'The stone is jade' in each language, and there must be other differences between the communities in order for the mental states of Ann and Jan to differ in truth conditions. The other differences will likely include intentional differences. If so, then I am free to exploit such differences as are required for Ann's and Jan's mental states to differ in truth condition. I need only claim that Ann's and Jan's local contexts are physically similar, not that there are no other differences elsewhere in the communities.²⁸

In sum, I can not find an interpretation of the cross-context test that blocks my counterexample without also ruling out uncontroversial relational properties (like being a planet) as unsuitable for scientific taxonomy. The moral is that broad contents are on a par with other relational properties, whose usefulness in science cannot be ruled out on a priori grounds.

BELIEFS AS STRUCTURING CAUSES

Beliefs are individuated by meaning or content, typically identified by 'that' clauses of attributions of belief in English. It is now generally agreed that semantic properties like meaning or content – properties in virtue of which states have truth conditions – are relational properties. If we assume, with the Standard View, that beliefs are internal states, this recognition poses the problem of showing how relational semantic properties of an internal state can be causally relevant to behavior. Content, or a state's having content, must help explain the behavior produced by the brain state. Otherwise, contents, and the beliefs that they individuate, are simply epiphenomenal.²⁹

The *prima facie* problem is this: According to the Standard View, meaning or content is a relational property of internal states,

28 Note that Ann's and Jan's communities are in the same possible world; so there is no question of whether their "whole worlds" are physical duplicates.

29 In "Anomalous Monism and the Problem of Explanatory Force," *Philosophical Review* 98 (1989): 153–88, Louise Antony has argued that Davidson's view of reasons as causes fails to account for the explanatory power of reasons.

but the causally efficacious properties of internal states are nonrelational. What actually causes one's finger to move (and the trigger to be pulled), for example, is intrinsic properties of brain states. If this is right, how can relational properties have any kind of causal or explanatory role in behavior? Fred Dretske tries to meet this challenge by showing how meaning, though a relational property of internal states, can still have a causal role in behavior. Dretske's account, like other naturalistic accounts of meaning for internal states, is two-tiered. At the ground level, meaning is linked directly to the (nonintentional, nonsemantic) physical world; then, with those naturalistic credentials in hand, other kinds of meaning may have social and linguistic components. Since I do not think Dretske (or anybody else) has furnished an "upper level" account of meaning, I focus solely on Dretske's ground-level account.

Dretske distinguishes between two kinds of causes: triggering causes and structuring causes. Suppose that a terrorist hooked up a car bomb to the ignition of a certain general's car; the general turned the key to start his car and thereby detonated the bomb. The triggering cause, which actually brought about the detonation of the bomb, was the general's turning the key in the ignition; but the structuring cause, which is responsible for there being a key-turn detonation process in the first place, was the terrorist's planting the bomb. The structuring cause (the terrorist's attaching the bomb to the ignition) may well be the causal explanation that authorities seek. The triggering cause sets in motion a process at a particular time; the structuring cause is whatever is causally responsible for the process to be in place.

Dretske finds a causal role for belief as structuring cause of behavior.³⁰ Suppose that an internal token of type *C* (where being of type *C* supervenes on intrinsic properties of the brain), causes a token of bodily motion type *M* on a certain occasion. Then the token of *C* is a triggering cause: Given the background conditions in which a $C \rightarrow M$ process is (in some sense) realized in the brain, the

30 Even if Dretske's account is successful in showing how meaning can have a structuring causal role in behavior, it still cannot explain any actual tokening of a behavioral process. Particular actions (such as shooting an intruder) would not be explained by showing how beliefs had a role in the shooter's being structured in such a way that when a certain internal event occurred, it caused a certain bodily movement.

token of C brought about the token of M . But the behavior has another kind of cause as well: The structuring cause is what brought it about that the $C \rightarrow M$ process is structured the way that it is.

Dretske builds up his account from the basic relation of indication.³¹ The idea is to give a nonintentional and nonsensical account of what an internal state means or represents (its content) in terms of what it indicates.³² Indication is a relation between token events: Token event b of type B indicates a token event a of type A if and only if (i) a caused b and (ii) there is reliable covariation between type- B events and type- A events. If a B -token indicates an A , then we may say that ' A ' is the natural meaning of the B -token. Obviously, indication or natural meaning is insufficient for representation. For there is no representation without the possibility of misrepresentation, and there is no possibility of "misindication." So, an indication theorist must move from indication (or natural meaning) to something that allows for error. Dretske does this by defining a new relation: *having the function of indicating something*. C may have the function of indicating F even if, on occasion, a token of C fails to indicate F – for example, the token of C is caused by something that is not F .

What is needed, and what Dretske supplies, is a naturalistic, ground-level account of how a natural indicator of F acquires the function of indicating F . Let C be a natural indicator of F . Then C acquires the function of indicating F if: (1) C is "recruited" (by a learning process) as a cause of M , where M is a bodily movement, and (2) C is so recruited because C indicates F . The recruitment structures a $C \rightarrow M$ process, so that, after the learning period, tokens of C cause tokens of M . Since it is because C indicates F that C is recruited as a cause of M , C 's indicating F is (in Dretske's terms) a structuring cause of the behavior M . C then has the function of

31 Fred Dretske, *Explaining Behavior: Reasons in a World of Causes* (Cambridge MA: MIT/Bradford, 1988). In *Explaining Behavior*, Dretske identifies behavior with the causal process $C \rightarrow M$; however, in "Mental Events As Structuring Causes," in *Mental Causation*, ed. John Heil and Alfred Mele (Oxford: Clarendon Press, 1993), 121–36, he applies his view to the more standard conception of behavior as simply the resulting M , not the whole causal process.

32 Although Dretske is (knowingly) casual about the type-token distinction, sometimes the distinction is important to his theory. For example, Dretske defines 'indication' as a relation between tokens; but his theory requires indication to be a relation between types. Talk of a state's structuring a $C \rightarrow M$ behavioral process likewise is to be understood as the state's structuring a type of process.

indicating F when the fact that C indicates F becomes a structuring cause of some behavior. Finally C 's meaning or representing F is understood in terms of C 's having the function to indicate F , where C acquired that function via a naturalistic learning process.

Now suppose that, after a $C \rightarrow M$ process is established by a learning process, a token of C occurs and produces the bodily motion M . Then, on Dretske's view, meaning has a causal role in the production of M , in virtue of the fact that C 's indicating F is a structuring cause of the $C \rightarrow M$ process. I believe that my remarks, though compact, accurately represent the structure of Dretske's view.³³ What I wish now to show is that it is thoroughly circular.

On the account just given, a state has meaning in the first place in virtue of its structuring a $C \rightarrow M$ process.

(a) A mental state C has meaning in virtue of its having a structuring causal role in a $C \rightarrow M$ behavioral process.

However, Dretske's goal is to show "how ordinary explanations, explanations couched in terms of an agent's reasons, explain."³⁴ Like other physicalists, Dretske takes explanatory role to be causal role. That is, Dretske's goal is to show how having meaning gives a state a structuring causal (and hence explanatory) role in behavior. If this is the goal, then Dretske is committed to the following:

(b) A mental state C has a structuring causal role in a $C \rightarrow M$ behavioral process in virtue of its having meaning.³⁵

(a) and (b) form a tight circle.³⁶ The circle is apparent in Dretske's characterization of beliefs as "those representations whose causal role in the production of output is determined by their mean-

33 I give a more detailed account in "Dretske on the Explanatory Role of Belief," *Philosophical Studies* 63 (1991): 99–111. See also Dretske's "How Beliefs Explain: A Reply to Baker," *ibid.*, 113–17.

34 Dretske, *Explaining Behavior*, 52.
35 For the moment, I am omitting quantifiers for ease of exposition. The intended reading of (a) is this: For any C , if C has meaning, there is a $C \rightarrow M$ process in which C has a structuring causal role, and in virtue of which C has meaning. The intended reading of (b) is this: For any C , if there is a $C \rightarrow M$ process in which C has a structuring causal role, then C has that role in virtue of having meaning. In the more technical discussion, I consider alternative readings.

36 It may be thought that Dretske is giving a logical analysis of 'having meaning', in which case (a) and (b) may be tautologous rather than circular. I believe, however, that Dretske aims to give an informative account of "the place of reasons in a world of causes."

ing or content – by the way they represent what they represent.”³⁷ If, as the account has it, meaning is itself determined by (structuring) causal role, then that same (structuring) causal role cannot in turn be determined by meaning.

Dretske objects that the account “isn’t circular because the causal roles meanings are supposed to explain aren’t the causal roles from which meanings are derived.”³⁸ Now both the causal roles that meanings are supposed to explain and the causal roles from which meanings are derived are structuring causal roles in behavioral processes. It may seem that Dretske is saying that there is no circle because C has two distinct causal roles in a single $C \rightarrow M$ process. But on Dretske’s theory, there is only one structuring causal role per behavioral process. Once the $C \rightarrow M$ process has been structured, a token of C becomes a triggering cause of the behavior; there is no more structuring to be done. Dretske says that C ’s meaning F is now a structuring cause of $C \rightarrow M$; but if C ’s meaning F is now a structuring cause of $C \rightarrow M$, it is so in virtue of the fact that C ’s *past* indicatings of F structured the $C \rightarrow M$ process: C ’s structuring causal contribution was completed in the past. So, there is no “logical space” for C ’s having two structuring causal roles.

Perhaps Dretske is distinguishing between two kinds of behavioral processes in which C has a structuring causal role: one behavioral process in virtue of which C ’s structuring causal role gives C meaning, and a different behavioral process in virtue of which C ’s having meaning is causally explanatory. That is, perhaps Dretske’s reply to my argument is this:

(Reply) For any behavioral process, $C \rightarrow M$, C ’s meaning F causally explains $C \rightarrow M$ if and only if there is some *other* behavioral process, $C \rightarrow N$, such that C -tokens’ past indicatings of F are the structuring cause of $C \rightarrow N$.

Of course, I agree that (a) and (b) do not really form a circle if a state C has meaning in virtue of its causal role in one behavioral process, but it has a causal role in some other kind of behavior in virtue of having meaning.

But (Reply) cannot be satisfactory. For it severs the behavior that C ’s meaning F causally explains from the behavior in which C had

a structuring causal role. In that case, the theory crumbles; for there is no provision for showing how C ’s meaning F has a causal role in *one* type of behavioral process in virtue of the fact that C ’s indicating F is a structuring cause of a *different* type of behavioral process. Unless the same behavior can be at issue in both (a) and (b), we would lose the explanatory link between the fact that C is a structuring cause and the fact that C ’s meaning has a causal role in behavior.

In the second place, (Reply) gives meaning an ubiquitous causal role: It follows from (Reply) that, for any state C that has meaning at all, the meaning of C causally explains all of C ’s behavioral effects. To put it another way, (Reply) has no room for a distinction between (i) cases in which C ’s meaning F causally explains behavior and (ii) cases in which C means F , but C ’s meaning does not causally explain the behavior C produces. For example, suppose that in the past, C ’s indicatings of danger structured certain avoidance behavior, so that C has come to mean ‘danger’. Suppose that on some occasion a C triggers some other behavior – say a blink of an eye, which is not structured by C ’s past indicatings of danger. According to (Reply), C ’s meaning danger causally explains the blink. This is intuitively wrong. Thus, I think that, although (Reply) avoids the circle, it undermines Dretske’s strategy for giving a causal role to meaning.

What Dretske is trying to show is that C ’s meaning something can give C a causal role in behavior. If meaning something were solely a matter of indicating something, there would be no problem: C would mean ‘ F ’ in virtue of the fact that C indicates F , and meaning ‘ F ’ would have a (structuring) causal role if C were recruited to cause M . But indicating F is not enough for meaning ‘ F ’. Meaning ‘ F ’ requires *having the function of indicating F* ; but by definition C does not have the function of indicating F unless C is already a structuring cause of some behavior.

So, here is the problem: Dretske takes meaning to be a structuring cause of behavior – whence the causal role of belief. C ’s meaning ‘ F ’ (at the ground level) is identified with C ’s having the function of indicating F . C ’s having the function of indicating F depends on the fact that C ’s indicating F is a structuring cause of the $C \rightarrow M$ process. Therefore, Dretske cannot – without circularity – take meaning, or the fact that a state has meaning, to be the structuring cause of $C \rightarrow M$ (or of the behavioral output M). For the structuring

³⁷ Dretske, *Explaining Behavior*, 52.

³⁸ Dretske, “How Beliefs Explain: A Reply to Baker,” 113.

of the $C \rightarrow M$ process is a precondition of C 's having meaning. The circularity slips in because meaning is implicitly identified both with C 's indicating F (meaning as structuring cause) and with C 's having the function of indicating F (meaning as representing F).

There is a way out of the circle, but only at the cost of giving up the explanatory or causal role of belief. The circle is generated, as we have seen, by a slide from the notion of C 's indicating F to the notion of C 's having the function of indicating F . It is only the former that is (noncircularly) a structuring cause of behavior; but it is only the latter that gives C a meaning: ' F ' becomes the (non-natural) meaning of C in virtue of C 's acquiring the function of indicating F .³⁹ So, the circle may be broken by consistently taking C 's indicating F (not C 's having the function of indicating F) both to be a structuring cause of behavior and to be a relation that *underlies* meaning or representation. But to say that a single relation is both a structuring cause and underlies meaning gives no causal role whatever to meaning. Meaning, on this noncircular rendition, remains wholly epiphenomenal.

Therefore, I believe that the most detailed attempt to provide an explanatory role for belief, construed as an internal state, does not succeed. For either it is circular or it accords the meaning of internal states no causal role at all. Although I cannot be sure, I believe that other naturalistic accounts of meaning, if developed in the detail of Dretske's, would fall to similar arguments.

39 In his recent "Mental Events As Structuring Causes," Dretske ignores the crucial distinction between indicating F and having the function of indicating F – noting only that a token of a type earlier tokens of which were indicators of F may misrepresent something as an F (135n). But the account of meaning requires the distinction between indicating F and having the function of indicating F , as Dretske acknowledges in *Explaining Behavior*. For example, assuming that tree rings indicate the age of a tree, suppose that bizarre weather conditions produced tree rings on a certain tree that did not correspond to age. Such tree rings would not be *in error*. They would not misrepresent the age of the tree in the sense relevant to misrepresenting a cow as a horse on a dark night. Even if there is generally a correlation between tree rings and the age of trees, failure of the correlation in a particular case would not be a mistake. In order for the unusual tree rings to misrepresent, they would have to have the *function of indicating* the age of the tree. Indication is not enough for misrepresentation; and as Dretske agrees, there is no representation without the possibility of misrepresentation. Ignoring the distinction between indicating F and having the function of indicating F obscures the circle, but does not eliminate it.

RELATIONAL PROPERTIES

One of the aims of this book is to give relational properties their metaphysical due. Often, it is in virtue of relational properties that something is the kind of thing that it is (e.g., a planet, or a husband).⁴⁰ And as many have argued, it is in virtue of relational properties that one has a belief that p . But what is a relational property?

We are accustomed to thinking of relational properties as expressed by two-place predicates of the form ' Rxy ' and thinking of nonrelational properties as expressed by one-place predicates of the form ' Fx '. But this is not the relevant contrast at all. For many one-place predicates express relational properties. Obvious cases (such as 'Sally is a sister') are easily converted into two-place predicates: 'Sally is a sister of x '. But other cases are less obvious: 'This bill is counterfeit'; 'Jones has tenure'; 'Michael Jordan was a basketball player'. It is unclear even how to express these relational properties as two-place predicates.

Let us understand relational properties broadly. Say that R is a relational property if and only if: x 's having R entails that there is some y distinct from x . In the language of possible worlds: R is a relational property if and only if for any world w and individual x , if x has R in w , then there is a y in w such that y is distinct from x . So, being counterfeit, having tenure, and being a basketball player are relational properties. There is a growing consensus, to which I am a party, that the property of having a belief that water is wet is likewise a relational property.⁴¹ For in a world in which S is the sole inhabitant (e.g., there is no water, or twater, or anything other than S), S does not have that belief. Predicates of the form 'believes that p ' express relational properties for any p which could not be an object of S 's belief in a world in which S is the sole inhabitant.

Say that a property P supervenes on local microstructure if and only if: Necessarily, if x has P and y lacks P , then there is a microphysical difference between x and y . Now relational properties expressed by predicates of the form ' Fx ' (such as 'has tenure' or 'is a

40 More radically – although I cannot argue for this here – sometimes it is in virtue of relational properties that something is the individual that it is.

41 Standard View accounts of belief in terms of relational properties are called "externalist" accounts.

planet') do not supervene on local microstructure. So, if one holds with Fodor that taxonomy in the sciences is by causal powers, and that relational properties can be taxonomic, then one is logically barred from taking causal powers to supervene on local microstructure. Suppose that Fodor is right – as I think that he is – when he says:

Taxonomy by relational properties is ubiquitous in the sciences, and it is not in dispute that properties like being a meteor or being a planet – properties which could, notice, distinguish molecularly identical chunks of rock – constitute causal powers. (MANC, 12)

Then, he should retract what in *Psychosemantics* he calls his “metaphysical point” about science: “Causal powers supervene on local microstructure.”⁴² However, abandoning that thesis in general would kick the motivation out from under the project of showing that a difference in broad content is not a difference in causal power in virtue of its responsibility for the properties of one’s behavior. For if causal powers generally do not supervene on local microstructure, why must mental causal powers supervene on local microstructure?

Here, I think, is Fodor’s rationale for holding mental causal powers to be locally supervenient: Mental causal powers are properties invoked by nonbasic laws. Nonbasic laws must be implemented by mechanisms that connect the satisfaction of the antecedents to the satisfaction of the consequents. In the case of psychological laws, the only plausible implementing mechanisms, claims Fodor, are neurological, and neurological properties supervene on local microstructure.

But to conclude from this that psychological properties must supervene on local microstructure is a non sequitur.⁴³ If we take implementing mechanisms to be chains of individual events, Fodor may be seen as claiming that for each sequence of individual events

⁴² Fodor, *Psychosemantics*, 44. In “Must Psychology Be Individualistic?” *Philosophical Review* 100 (1991): 179–204, Frances Egan has argued that this metaphysical point is false.

⁴³ Fodor may now agree with this point. In his post-*Psychosemantics* writing, he has placed increasing emphasis on implementing mechanisms and has not coupled his “implementing mechanism” thesis with the claim that psychological states supervene on neural states. See Chapter 6 for further discussion and criticism of Fodor’s claim that psychological laws must be implemented by computational mechanisms that supervene on neural processes.

subsumed by a psychological law, there is a sequence of individual events subsumed by neurological laws. But, as Burge has argued, this claim entails nothing about the individuation of event types.⁴⁴ For the following are consistent (whether true or not): (1) Neurological properties supervene on local microstructure; (2) psychological laws are “implemented” by neurological mechanisms; (3) properties projected by psychological laws do not supervene on neurological properties; nor do psychological property instantiations supervene on neurological property instantiations. From the fact (if it is a fact) that neurological mechanisms “implement” psychological laws, it does not follow that the neurological properties of any individual event fix the psychological properties of any individual event. Psychological properties may fail to supervene on neurological properties because individuation of psychological states is more sensitive to the subject’s environment than is individuation of neurological states.

Indeed, in general, properties of a higher-level process do not supervene on properties of mechanisms that implement the higher-level process. To take a commonsense example, consider a presidential press conference carried live on television. The political property of being a televised presidential press conference does not supervene on the intrinsic properties of implementing mechanisms. A microphysically duplicate mechanism may implement something quite different from a presidential press conference. Or consider the mechanism by which the automatic-teller machine gives me money from my checking account. The same mechanism could be used to implement an entirely different process; it could, for instance, give me green pieces of paper (functionally equivalent to “pink slips”) whose numbers tell me which employees to lay off. There are endless examples like these. The moral is that the intrinsic properties implementing mechanisms do not generally fix the properties of the processes that they implement. So, the fact (if it is a fact) that neural mechanisms implement psychological laws would provide no motivation for narrow taxonomy in psychology.

⁴⁴ This general line of thought is advanced by Tyler Burge, “Individualism and Psychology,” *Philosophical Review* 95 (1986): 3–46. A similar argument is developed in “Individuation and Causation in Psychology,” *Pacific Philosophical Quarterly* 70 (1989): 303–22.

CONCLUSION

To begin to dislodge the Standard View, I have presented three attempts to work out a Standard View conception of belief that would give beliefs a causal-explanatory role in behavior. All three, I have tried to show, come to grief. Although my arguments do not prove that no Standard View account of belief can ever succeed, they should give pause: The range of views discussed is broad, and problems encountered are deep. Alongside those proponents of the Standard View who undertake to show how beliefs can be constituted by particular brain states are eliminative materialists, who, reading neuroscience as casting doubt on the thesis that beliefs are constituted by particular brain states, give up on belief altogether. After trying to cut the ground out from under eliminative materialism in the next chapter, I argue in Part II that the causal explanatory-ness of belief does not require that they be identical with or constituted by particular brain states at all.

The m

Proponents of the eliminativists, who beliefs, and eliminativ be coherently identical there are beliefs or beliefs, eliminativist The arguments for e construing common dubbed 'folk psychology. Although I do the scientific psychology seriously mistaken psychology: It is not The arguments for ise with which I am premise

(S) Beliefs and not be vindic the eliminative mat

(EM) There are r
The truth of (S)
program in psychol
beliefs are syntacti
were supplanted by

¹ The literature on
sense: (1) a sense
weaker reading
"modified"
Critique of
here, see p.