

20 Epiphenomenalism: Reply to Dretske

Fred Dretske's essay presents a tantalizing and engaging combination of agreement and difference. We agree on anti-individualism, or what he and others call "externalism." We agree on realism about intentional content, on the real existence of intentional states and events, and on the explanatory and causal relevance of such states and events, individuated nonindividually. I agree with most of his invocations of the relevance of function in intentional explanations. And I see his application of intentional explanation to animals that lack language, and his discussions of artifacts like gauges and computers, as congenial and illuminating. We differ over whether materialism is obviously true and whether naturalistic reductionism is a worthwhile enterprise. More to the point of the main lines of his essay, we differ about whether epiphenomenalism is a serious threat and how to go about showing its impotence.

This difference over the status of a threat from epiphenomenalism is related to a lesser difference over sociology. Dretske writes that he thinks that he and I may be the only ones who believe that the content of intentional states can be both externally individuated and causally relevant to the explanation of behavior. I enjoy and value Fred's company, but I do not feel nearly so isolated or beleaguered as he does. I think that, despite the widespread discussion of epiphenomenalism and the posing of "worries" about it, nearly everyone except for the occasional wide-eyed revisionist believes that intentional explanation is in part causal explanation and is relevant to the explanation of mental events and behavior. And I think that anti-individualism has carried the day among the vast majority of philosophers who have understood and reflected on it. The intersection of these groups constitutes a large majority of philosophers who have thought about these subjects.

Of course the sociology of philosophy is of no great importance in the short run. But I believe both of these positions will prove very stable. This judgment is the historical analogue of my view that much of the recent discussion of epiphenomenalism, in the light of anti-individualism, is conducted in an atmosphere of make-believe—a mock phobia that is dramatized mainly to make some metaphysical model seem like desperately needed salvation from impending psychic or philosophical disaster.

My view is *not* that philosophical issues about mental causation or mind-body causation are uninteresting or pointless. Dretske's and others' out-of-context citations sometimes make it appear that that is my view. But it isn't, nor have I ever expressed such a view. There are difficult and worthwhile issues in this area. My negative attitude toward the ways these issues have been discussed has two components. The first concerns the antecedent make-believe posture toward the discussion, that treats epiphenomenalism as a serious threat rather than as a test or limit position that any sound view must avoid. The second concerns prevalent assumptions—centering on materialist metaphysics—about what sorts of considerations must be dominant in an explanation of why epiphenomenalism is false.

Let me begin with the point about make-believe. Dretske's comparison of the problem to Zeno's paradoxes seems to me to be in the right spirit. No serious person, even a philosopher, thinks that we cannot traverse a space. But there can be philosophical interest in explaining what is wrong with Zeno's reasoning that says that we cannot. That is the spirit in which epiphenomenalist doubts should be treated. I think that despite his wonderfully relaxed philosophical manner, Dretske's rhetoric of "worry" and "epiphobia" that needs to be "quieted," seems to me to lose touch with the Zeno paradigm. I think that we know that beliefs and desires are causally efficacious as such with as approximately much certainty as we know that we can traverse a space. Even granting the materialist-reductionist starting point that Dretske assumes, it seems to me that epiphenomenalism can be safely presumed to be false. The discussion should be pursued in that spirit, not in a spirit of angst, which I cannot help but find artificial and histrionic.

Dretske produces quite a range of examples that are supposed to raise the epiphenomenalist worry. But taken as potential threats, none of them seems to me to be intellectually powerful. Relevant disanalogies with the mental state/event cases are in each instance immediately obvious. At most, the examples prod one to think more clearly about mental causation. Contrary to what Dretske suggests, we do not disagree about the causal relevance of the property of being an aunt or uncle. The comparison with the relational property of being an uncle suggests only that some relational properties are not causally relevant, not that all relational properties (much less all nonrelational properties whose nature presupposes the existence of relations) are causally irrelevant. Causally relevant relational properties are abundantly invoked in the natural sciences. No specific reason is given in this example for thinking that mental properties are not causally relevant. We have overwhelming grounds to presume that they are.

Similarly, no one thinks that being a twenty dollar bill or being an H-particle or being a quarter is causally relevant to the transactions that are explained in terms of underlying physical properties that these properties happen to be correlated with. Such examples show at most (see below) that some higher-level properties are causally irrelevant, not that all are. Even assuming that all mental properties and events are correlated with specific underlying physical properties, which is highly doubtful, nothing specific to mentalistic causation has been cited that would suggest any error in the common presumption that mental states/events as such are causally relevant in a wide variety of causal transactions. Citing these cases is somewhat like saying that since a number cannot traverse a space, maybe physical bodies cannot either.

Dretske notes that on the anti-individualist view it is assumed that at some level of description the twins, in the standard thought experiments, will be undergoing and even doing some things that are the same. One can use this as a ground to believe that anti-individualist explanation is causally irrelevant only if at least two conditions are met: First,

explanations of the ways, particularly intentional and function ways, in which the twins are *different* are causally irrelevant. Second, where the twins do the *same* things, there can be no differences in mental causes due to differences in the causal structures and sequences in the two environments. I see no reason to think that either of these conditions is met.

Dretske lays the first condition aside in his motivational section, without explaining why he does so. He does not discuss the second. But one can raise one's arm for different reasons. It is not clear why the causes of raising twin arms in the different environments cannot also differ, given that the arm is raised for some purpose, whose eventual effects will have different properties.

Assuming that these two conditions obtain really amounts to assuming that anti-individualism about mental states/events is incompatible with the causal relevance of intentional aspects of those events, which is exactly the point at issue.

Dretske's most specific formulation of his worry is as follows: "The worry is not that externally individuated beliefs do not play a role in the explanation of behavior. . . . The worry is that their role is dispensable—that the same explanatory results can always be achieved without mentioning those aspects of belief that fail to supervene on the believer" (p. 156). I think that this "worry" has only to be stated and reflected on to see how wildly implausible it is. No one has made a remotely persuasive case that nonmentalistic explanation, or some artificial form of explanation that postulates only "narrow" content, can achieve *all the same* explanatory results that mentalistic explanation does. Although in some particular cases, one can hive off particular external aspects of belief (e.g., the *re* in *de re* belief), no one has even begun to provide an explanatory replacement for the basic intentional explanatory elements in mentalistic explanation. Those elements and many of their *explananda* are individuated nonindividually. Nor is there strong reason to think that intentional, mentalistic explanation is, throughout, noncausal.

Dretske has a tendency, in expounding the "worry," to collapse the distinction between the mental states/events as individuated by their content—which are the explanatorily relevant properties—and the relational facts that are necessary to the having of that content—which are normally not themselves part of a causal explanation, or causally efficacious, at all. If this point is kept clearly in mind, the worry about the wholesale dispensability of mentalistic explanation seems to me vanishingly thin.

There remains worthwhile inquiry into the specific ways that mentalistic causal explanation differs from underlying natural-scientific explanation, which prevent mentalistic causal explanation from being dispensable. Dretske has made substantial contributions to this inquiry. I think that his distinction between behavior, as a process consisting of a sequence of events, and motion is insightful and relevant. And I think that his appeal in his (1988) to structuring causes, though not in my view a full account of the causal relevance of reasons, is also a significant contribution. It is notable that despite the

materialist cast of his account, neither of these contributions depends essentially on his materialist framework.

Even apart from Dretske's particular account of behavior as a sequence or process, mentalistic explanation of simple events like raising an arm, which are not specified in intentional terms, is normally explanation of an act, not a mere happening. Even if the doing is to be counted identical, sometimes, with a happening, the aspect of the event that is explained by mentalistic explanation—even when it is not an aspect directly involving intentional content—is different from the aspect of the event (mere motion) that is explained by a physiological or purely physical explanation. Further, as we agree, often it is the intentional content of behavior that is explained. These reasons are just a small sampling of obstacles for physical-causal explanation's replacing mentalistic explanation.

Dretske makes much of a need to provide a metaphysical account of causation in light of the supposed unreliability of explanatory practice in signaling the causal relevance of mental properties. I do not deny that metaphysical models are worth having when we can get them. But I think that his motivation of a need for them by denigrating the role of explanatory practice in guiding us to the causal structures of the world is misconceived. It depends on too loose a conception of explanatory practice. Of course, there are scattered rule-of-thumb explanations in common sense that can be recognized on a moment's reflection not to pick out the causally relevant properties, at least relative to certain explanatory enterprises. But I have in mind more fundamental aspects of explanatory practice in everyday life and in psychology.

Consider the widely used Coke machine example. Being a quarter is epiphenomenal relative to the mechanical causation of a Coke machine. Recognition of this point is surely part of even moderately reflective aspects of explanatory practice. I think, however, that certain explanations of functional aspects of the "doings" of even a Coke machine can be causally relevant, along the lines of Dretske's structuring causes. I agree with him that the functions of the machine are not autonomous and certainly not mentalistic in this case. I agree that the case is not a good model for the causal elements in mentalistic explanation. The issue of whether an object's being a quarter is a causally relevant property does not seem to me very important.

I am not convinced, however, that the property of being a quarter is properly seen as epiphenomenal with respect to *all* causal transactions. Suppose that we allow substitutes—counterfeit slugs—to count as having the causally relevant properties in particular cases, as most causal explanations of the functional aspects of events do. Such substitutes would succeed in causing particular instantiations of functional properties via other nonmonetary causally relevant properties. The slug would cause the Coke machine to deliver a Coke and change. The substitutes would succeed in explaining, relative to background assumptions about the functional relation, the same instantiations of functional properties as the

“noncounterfeit” causes would. But if the substitutions were made wholesale from the beginning in all cases, the substitutes would not succeed in causing the functional aspects of the *explanandum*. If there were no normal causes—instances of noncounterfeit input coins—there would be no Coke machines yielding Cokes and change. One can explain any given motion of the machine without appealing to the property of being a quarter. But one cannot fully explain the functional properties of the motions—the fact that the machine provides a Coke and *change*—without appealing to monetary properties. This seems to me already to differentiate the Coke case from the standard epiphenomenalist model. But as noted, I agree with Dretske that the Coke machine example would be a poor model for mental causation.

Dretske’s examples of a \$20 bill’s effect on the cashier’s visual system are similar. It is easy to convince ourselves that the visual properties of the \$20 bill, not its monetary properties, are what are causally relevant to purely physicalistic aspects of the causal transaction with the cashier’s visual system. The purely physically described transactions focused on do not need explanation through the higher-level properties. The explanations for higher-level properties that we are commonly invited to consider are offhand, rule-of-thumb explanations that are not embedded in a systematic explanatory system. It is easy to imagine dispensing with such explanations, at least relative to the aspects of the *explananda* that Dretske has in mind. But even in these cases, it is far from obvious that all causal explanation can dispense with the higher-level properties.

To fully understand the situation, one needs to think about what the account of vision is supposed to do. What are the causally relevant properties? In one sense, anything that causes the same retinal stimulations. Any particular instance of a visual representation can, in a sense, be causally explained as well through appealing to a counterfeit bill, or a quantum-accident look-alike, as by the \$20 bill. Considering only such cases would be to follow the method of Dretske’s argument “anything like the \$20 bill that was sufficiently similar would have the same effects.” But we want to specify the cause as what is seen too. Which properties of what is seen are causally relevant? At a certain level, just the “visual” properties—properties of color, shape, texture, shading, distance, and so on. But there are perceptual concepts that derive from training. These include more than visual properties in the most psychologically conservative sense. People visually recognize horses, cancer, money. To account for the representation of these properties, we need to cite the properties at a certain level in the explanation of visual perception. Other look-alike properties would serve in the causal explanation of any given (false) visual representation of normally seen properties. But look-alike properties could not suffice to explain the acquisition of visual information about the original properties. Nor would they explain the function of the visual system, or higher-level representational systems, in representing the original properties. A systematic causal explanation of *even individual instances*

of perceptual representations will require appeal to higher-level properties in the generalizations, as normal causes.

Moreover, mentalistic explanation that makes use of mental events as causes must be able to explain particular events. Dretske's own account of mentalistic explanation pays too little attention, in my opinion, to the role of mentalistic properties in causal relations between particular events. Particular acts and events are caused by events with mental properties—sensations, perceptions, decisions, judgments. We have every empirical reason to think that such explanation is not only genuine but irreducible and ineliminable.

Even the initial points made with respect to the Coke machine and cashier cases fail to apply in the case of mentalistic properties generally. We have no purely physical explanations of our detailed *physical* movements, much less of our intentional activity and behavior. We do not even have such explanations for the activity of the higher animals. We have massive empirical reason to believe that psychological causal explanation, with appeal to causal efficacy of instances of psychological kinds, is indispensable for explaining a wide range of important phenomena. A satisfying metaphysical model of mental causation would be welcome. But no such model is needed to show that such causation is real.

An undercurrent difference between Dretske's approach and mine lies in our attitudes toward the epistemic status of materialism. He begins with a materialist metaphysical model, assuming that as known. He holds that model fixed in trying to explain why epiphenomenalism is not true. I have indicated that I think that epiphenomenalism is no serious threat even *given* this starting point.

I have no objections to metaphysical inquiry in this area. Where we can get deeper understanding from metaphysics, we should, of course, get it. I do not, however, accept as obvious or fundamental the assumption of the materialist and mechanist model that Dretske starts with. I think that whether materialism is a correct metaphysical view of ourselves and the mind depends on among other things *exactly* what the relations are between thoughts and the underlying physical material. The mere fact of some necessary dependency will not suffice for materialism. The relations must be identity, composition, or something similar. I think that we do not have satisfying accounts of exactly what these mind-body relations really are—in particular whether they are forms of material constitution.

Just maintaining that mental events and properties need a material substrate—or need to supervene on such a substrate—does not suffice to support materialism. For materialism is committed to specific relations—identity or material-composition relations—between the mental items and the physical items. Assuming a metaphysical model with these commitments seems to me to go far beyond anything that our actual explanatory

practice warrants. In fact, it seems to me that materialist metaphysics has succeeded in obscuring mind–body causation more than it has illuminated it.

I have not been persuaded by the blizzard of reasons given to insist that mental events are material events. All the materialist accounts of phenomenological events and even of propositional attitude events seem to me to have a whistling-in-the-dark character. Materialist metaphysics will be and should be an ongoing enterprise. But it should be taken as experimental and conjectural, not as the philosophical starting point that it became for so many philosophers in the second half of the twentieth century. No particular form of materialism has earned a status of orthodoxy. And a generalized commitment to materialism seems to me to owe more to pious hope, wild inductive projections, and caricatures of nonmaterialist alternatives, than to rational argument or empirical support. I believe that the next century is ripe for fresh ideas on the mind–body problem.

References

Dretske, Fred. 1988. *Explaining Behavior*. Cambridge, MA: MIT Press/A Bradford Book.