[*Acta Analytica* 11 (1993): 111-124]

Dretske on Explaining Behavior

Kirk A. Ludwig
Department of Philosophy
University of Florida
Gainesville, FL 32611-8545

1

Many philosophers have recently been concerned with whether our ordinary practices of explaining what we do in terms of reasons can survive the discovery of the neurophysiological causes of our behavior, conceived of as the movements of our bodies through space. For example, when I sat down to write this paper, I raised my arm to turn on my computer. Why did I my arm move? Surely there is a neurophysiological explanation of the motion of my arm, even if we don't yet know how to give it. But if there is, then what work is there left to do for my desire to understand how my reasons explain what I do, my intention to write a paper, and my belief that if I am going to do it, I'd better sit down and turn on my computer? The worry is that there cannot be two explanations, two causal explanations, of the motion of my arm, and that, therefore, reason explanations must be given up, since, in this conflict, neurophysiological explanations are sure to win.

This worry, I think, is misplaced, but I do not want to address it here, but rather a prior question, namely, whether action explanations are so much as in the same line of business as neurophysiological explanations of the motions of our bodies and limbs. If they are not, then the existence of neurophysiological explanations for our movements is no threat to the legitimacy and explanatory power of action explanations. What gives rise to the thought that they might not be in the same line of business is that what is explained, when we give a neurophysiological explanation, is (a) *the rising of my arm,* while what is explained when we give an action explanation is (b) *my raising my arm.* And while the latter entails the former — if I raise my arm, my arm rises — the former does not entail the latter, since my arm may rise without my raising it. This difference forms the starting point of Fred Dretske's recent and highly original

response to this worry in his book *Explaining Behavior* (Dretske 1988).[1] I think Dretske is right to draw attention to the distinctive form of the explanandum in action explanations, which plays a central role in his defense. But I think the moral he draws from it is incorrect. The burden of this paper is to show why Dretske's account, which I think is the most promising account so far of how to sidestep the problem of the explanatory exclusion of psychology by neurophysiology, cannot be right. The failure of Drestke's account, and in particular the reasons it fails, should force us to re-examine the assumptions which generate the problem of the explanatory exclusion of psychology. I begin with a brief sketch of Dretske's account.

2

Dretske's account of the explanatory role of reasons rests on two important claims about our explanations of behavior (Dretske 1988, chapters 1 & 2). The first is that what we explain, behavior, is a *process,* the process of some internal state of type C *causing* some bodily movement of type M, and then further processes which extend beyond the body and have this process as their initial segment. (Below I will use capital letters to indicate state types or tokens of a state of a certain type; the context will make clear which.) This is motivated by the special form that the explanada of action explanations take. The explanandum for an action explanation is of the form 'A φ-d' where 'A' is a singular term that refers to an agent, and 'φ-d' is replaced typically by an action verb, which is often followed by a direct object. Why did I *move* my arm? Why did I *lift* the cup? Contrast these requests with: Why did my arm move? Why did the cup rise? The latter are clearly asking for an explanation for the occurrence of an event. The former, instead, seem to be asking for an explanation for a causal process, my *moving* my arm, or my *lifting* the cup.

The second important claim is that action explanations are what Dretske calls structuring cause explanations, rather than triggering cause explanations. This distinction between kinds of explanation relies on a distinction between kinds of causes, between *triggering causes* and what Dretske calls *structuring causes*. An example will help to make clear this distinction. The thermostat in my house turns on the heater when the temperature drops below 68 degrees

Fahrenheit. The temperature's falling below 68 degrees Fahrenheit is the triggering cause of the thermostat's turning the heater on in my house. It triggers a certain process — as we might generously put it, a bit of behavior. But what caused the thermostat to be so arranged as to turn on the heater, rather than the washing machine, is not the drop in temperature, but rather a certain person's movements some time in the past. This is what Drestke refers to as the structuring cause, what caused this system to be so *structured* that when the temperature drops below 68 degrees, it turns on the heater rather than something else, or nothing. If C is the internal state of the thermostat when the temperature drops below 68 degrees, and M is its turning on the heater, this is what accounts for C's causing M, rather than something else. Note that the structuring cause is a cause of the same thing that the triggering cause is a cause of. And it is in its own right a triggering cause of the condition of the thermostat being hooked up to the heater. Thus, the concept of a structuring cause must be relativized to an effect (type) and a background condition necessary for the effect type to occur given its proximal triggering cause.[2] The structuring cause thus causes the effect by being a cause of a condition necessary for it. A structuring cause explanation is an explanation of why a certain event brings about another which cites a structuring cause under a causally relevant feature. Citing the structuring cause then explains why the first event brings about the second via an explanation of why the system is structured so that it does so. If action explanations are, as Drestke claims, structuring cause explanations, this would help to explain the special form of the explananda in action explanations.

The relevance of this to giving *reasons* a role in the explanation of behavior is provided by Dretske's account of how an internal state of a system acquires a representational content. Dretske's account here builds on and modifies the account he began in *Knowledge and the Flow of Information* (Dretske 1981). Central to this is the notion of a natural sign of something, an indicator. S indicates that *s* is F just in case the probability that *s* is F given S (and certain fixed background conditions) is one.[3] Since a sign indicates that *s* is F only if *s* is F, indication is not representation, for whatever represents can misrepresent. An internal state *represents* that *s* is F when it acquires the function to indicate that *s* is F. It acquires the function to indicate that *s* is F when it is recruited to cause some particular output of the system *because* it indicates that *s* is F (Dretske 1988, chapter 4). "[B]eliefs are precisely those internal structures that have acquired

control over output, and hence become relevant to the explanation of system behavior, in virtue of what they, when performing satisfactorily, indicate about external conditions" (Dretske 1988, 84). Such states then "exhibit the essential properties of genuine beliefs: they *have* a propositional content, and their possession of this content helps explain why the system in which they occur behaves the way it does" (Dretske 1988, 107). Thus, according to Dretske, reasons, beliefs in particular, help explain behavior, conceived of as C's causing M, because C, the internal state that is the reason, acquires its content (becomes a reason) in virtue of being recruited to cause M.

While this is the centerpiece of Dretske's proposal, it is not the whole of it, for we still need to say where desires fit into this picture. Desires come in in the following way. Dretske identifies the tendency for a link to be built between an indicator and M, when M results in a certain stimulus R when the indicator occurs, with a *pure* desire for R (Dretske 1988, chapter 5). Thus, the content of the desire, the fact that it is a desire *for* R, is also supposed to play a role in explaining why C, some internal structure, comes to cause M, and so to provide part of the structuring cause explanation of C's causing M, and so to complement the account of the explanatory role of beliefs.

3

This account has a number of virtues. It explains the special form of the explanandum of action explanations. It solves the problem of explaining how reasons can have explanatory work to do if there are adequate neurophysiological triggering cause explanations for our movements. It seems to avoid any difficulties there might be thought to be in giving reasons an explanatory role if we think that their contents must be relationally determined, and that only non-relational features of an individual's body could be causally relevant to the motions of his limbs. And it provides a novel naturalized account of representation as an integral part of the story. Nonetheless, it cannot, I think, be right.

Broadly speaking, there are two jobs that action explanations do. The first is to explain why *this* bit of behavior occurred rather than some other or none, and so to provide a causal

explanation of the behavior. The second is to explain what minimally was to be said in favor of the behavior from the point of view of the agent — in Davidson's (1963) phrase, they rationalize the action. It is not clear that Dretske's account provides adequately for the rationalizing role of reasons in action explanation (Stampe 1990). But my objection will be rather to Drestke's account of how reasons provide a *causal* explanation of behavior. I will also limit my attention to Dretske's account of belief. Criticisms similar to those I give of Dretske's account of belief can be given for his account of desire. Belief and desire are equal partners in the explanation of action, and their role should be parallel. In any case, if Dretske's account fails for either, it fails as an account of how reasons can causally explain our behavior.

Before considering structural problems with Drestke's account, I want to begin with an intuition. Dretske's account requires requests for action explanations to be requests for facts about our learning histories. This strikes me as on the face of it an implausible claim. Even cursory reflection on what we are interested in when we ask for explanations of the behavior of people around us should convince us that we are not interested in facts about their distant past, as we would be on Dretske's view. To see this, suppose that you were convinced that someone (perhaps yourself) had been in existence for exactly five minutes.[4] We would *still* feel perfectly comfortable in explaining his behavior by citing his reasons; this is especially clear if one considers it in the first person case. But on Dretske's view we could not possibly do so. Since, it seems to me, we obviously *can,* I think we would be justified on this ground alone in concluding that ordinary action explanations are not structuring cause explanations.

But I do not want to rest my criticism of Dretske's position on this intuition. I turn now to two structural problems with Dretske's account, which, I think, show that it cannot possibly be correct.

(1) The first difficulty is raised by what I will call the *plasticity* of behavior. This is the familiar point that beliefs have behavioral consequences only as embedded in larger networks of beliefs and desires. Suppose that I believe that you have insulted me. What behavior does this cause? If I am benign or timid, it may cause none at all. If I am impetuous, and easily angered, it may cause me to return the insult, or to strike you, or to begin cunning plans to bring about your downfall or public embarrassment. The list of possible consequences goes on and on, so it is not

possible simply to settle for a disjunction of these. The list is open-ended, and depends on my other beliefs and desires, habits, attitudes, and dispositions, all of which are to some degree in constant flux. If I behave in a certain way now, I may behave in a different way next week when I again come to believe that you have insulted me. So if you are told merely that I believe that you have insulted me, you are not thereby in a position to predict anything about my behavior.

This raises the following difficulty for Dretske's account. In the account given in chapter 4 of *Explaining Behavior,* the content of a belief becomes relevant to the explanation of the process of C's causing M because it is acquired in the process of the system's coming to be so arranged that its internal state C causes M. Thus, what, if anything, the reason aids us in explaining, on Dretske's account, is the process of C's causing *M*. If C later causes M', citing the content of C will be utterly irrelevant to explaining why this process occurs in the system. Thus, the account must leave out most of the actual and potential behavior of which on Dretske's account the belief is a part. Consequently, the work that citing a belief with a certain content typically does in explaining behavior cannot be to explain why some internal structure of that type, which happens to be a belief, regularly causes behavior of the type actually caused.

The objection here is not that there are many different ways to engage in a given kind of behavior, e.g., tying one's shoes. It is that no matter what the level at which one types a kind of behavior, a belief will not cause that kind of behavior except as it is embedded in a larger network of beliefs and desires. The only way to define output so as to *guarantee* that the same output was always produced by a belief with a given content (when anything was) would be to define the output type as something caused by a belief with that content. But this would make nonsense of the idea that beliefs are internal structures recruited to cause a particular type of behavior because of what they indicate, so this is not an option open to Dretske.

It might be said on Drestke's behalf that, although this account won't work, a more sophisticated account on the same general lines will. What Dretske needs to say is that a belief gets its content not by being recruited to a cause a particular bodily movement, or even some bodily movement, but to play a certain *causal role* in producing the believer's behavior. Dretske must say that the internal state C is recruited to cause *various* behaviors, which behaviors depending on what other states are present in the system at the time. Then, when we cite a belief

in the explanation of some novel behavior, we will be explaining not why the organism is so structured that that internal state causes bodily movements of this type, but why the organism is so structured that it causes bodily movements of this type when it is conjoined with such and such other states.

The trouble with this modification is that simple reinforcement can no longer be a model for how an internal state is recruited to play the appropriate role, but it is not clear that there is any coherent substitute. Desires cannot, on this new picture, be simply the tendency to build a link between the internal state C and the bodily movement M, where M produces a reinforcing stimulus in the learning situation. The desire must be thought of as the tendency for the internal state to acquire a certain *causal role* in the system, which is not tied to the production the movements it actually produces in the learning period. What is puzzling about this revised picture is how the specific content that is assigned to each belief could be recruited play the appropriate coordinating role in producing behavior. For that role is one in which what the state causes depends on its specific content and the content of other states in the system. But the tendency to be recruited for a certain role is identified as a desire, which is a non-relational state of the system. That depositional state of the system is sensitive only to proximal stimuli. Certain proximal stimuli will count as reinforcers. But they will do so independently of what distal state of affairs is indicated by the state being recruited to play a certain role in the system. How then can it be sensitive to relations between the system and its environment in the way that would be required to give the particular content its *special* role, a role that must take into account the distal events or states of affair it actually indicates in the learning period?

(2) I doubt that the problem of the plasticity of behavior can be adequately dealt with in the framework that Drestke provides for us. But now I want to turn to a deeper difficulty. Even if we were to waive this objection, given *how* Dretske says an internal structure acquires the function to indicate that *s* is F, the function to indicate that *s* is F could not play the role it has to in order to provide a structuring cause explanation of our behavior.

To see why, we can first note that the content of a belief is not itself a causally relevant feature of the structuring cause. A structuring cause explains why an internal state of type C comes to cause a bodily movement of type M. On Dretske's account, when C comes to have a

representational content, and so comes to have an explanatory role as a reason, what explains this is a certain property of the state type C, namely, that it indicates, e.g., that *s* is F. If C is recruited to cause M because it indicates that *s* is F, *then* (and only then) C has the function to indicate that *s* is F. But this means that when C has acquired the function to indicate that *s* is F, C *already* causes M. Hence, C's having the *function* to indicate that *s* is F could not explain why C causes M. What explains C's causing M is not that C has the function to indicate that *s* is F, but C's previously *indicating* that *s* is F. So it is not the representational properties of C that explain why it causes M, but that in the period when it was recruited to cause M, it indicated that *s* is F. The representational properties of C, according to Dretske, are historical properties of it, properties it has in virtue of having had a certain history. But if we want an explanation of C's coming to cause M, we cannot offer the fact that it has come to cause M or some bodily movement, even if we add that it has come to do so for a certain reason. Therefore, even if reason explanations were requests for structuring causes, given Dretske's account of how states acquire representational content, reasons could not fulfill that role.

One response to this objection is to say that, while representational properties are not properties of the structuring cause, and, hence, not causally relevant to C's coming to cause M, they nonetheless are explanatorily relevant, because they function as pointers to what are the genuinely causally relevant features of the structuring cause.[5] This strikes me as implausible. Whatever its merits, though, we can note that it is clear from this that reasons are neither structuring causes nor causally relevant to behavior however conceived. So if Dretske's project is to show how reasons can be causally relevant to behavior, as I believe it is and should be, this account fails to show that. Apart from this, what is puzzling about this response is that if this is the only role that reason properties play in explaining behavior, then it is mysterious why we should appeal to them. Why do we not simply advert to the past informational properties of C?

This last question points to a way of reformulating the objection that brings out more clearly the underlying structural problem with Dretske's account. It *is* a problem for Dretske's account that representational properties are not causally relevant features of structuring causes. But when we look more closely at the reason this is so, we can see that given the way Dretske's account assigns content to reasons, the content could not be in any way even *explanatorily*

relevant.

In ordinary action explanations, we cite reasons in the explanans. For example:

A: Why did you open the window?
B: I opened the window because I wanted some fresh air and thought that opening the window would help.

This requires that the equivalence Dretske gives us provide an explanation of my behavior when substituted into the context following `because'. When we pay attention to this requirement, we find that Dretske's account is faced with a dilemma, neither horn of which can be acceptable to him. The dilemma is that either Dretske's analysis of representational content has the wrong form to play the kind of explanatory role he gives to it or the account assigns representational content to everything.

According to Dretske's account,

> C represents that *s* is F iff *C was recruited to cause some bodily movement* because *it indicated that s is F*.

Suppose we request a structuring explanation of C's causing M, that is, we ask, Why does C cause M? An answer to this question will take the form:

> C causes M *because* ...

where `...' is replaced by the explanans. On Dretske's account, we explain why C causes M by citing its representational properties, thus,

> C causes M *because* C represents that *s* is F.

Now let us substitute in for `C represents that *s* is F' using our equivalence above to get,

C causes M *because* C was recruited to cause some bodily movement *because* it indicated that *s* is F.

Before proceeding, let us ask what 'C was recruited to cause some bodily movement' means. The use of the word 'recruited' suggests that we are already alluding to the explanation of C's coming to cause some bodily movement. But since we are here giving that explanation, we should not build any of it into the explanandum, and so we should read this simply as 'C came to cause some bodily movement'. However, this cannot be quite right, since we want our explanans for 'C causes M' at least to entail it; but the fact that something explains why C came to cause M in the past does not entail that C now causes M. So we should read the embedded explanandum above as 'C came to cause and now causes some bodily movement'. We can shorten this to 'C causes some bodily movement' since saying 'C causes some bodily movement because C indicated *s* is F' entails that C came to cause some bodily movement. Thus, the canonical form of the proposal should read:

C causes M *because* C causes some bodily movement *because* it indicated that *s* is F.

Two points are apparent from this which we have already noted. First, the fact that C causes some bodily movement for some reason could not in itself explain why it now causes M, for the sufficient reason that we do not know whether the bodily movement C came to cause is M or some other bodily movement type. Second, the plasticity of behavior guarantees that for many bodily movements which C causes our explanans could not be helpful.

But now we can see another problem with the proposal. We are offering as an explanans for 'C causes M' an expression that is itself a statement of the fact that one thing explains another, and in particular a statement that a certain thing explains why C causes some bodily movement. And the explanandum here, 'C causes some bodily movement', is entailed by our original explanandum, 'C causes M'. Now intuitively the difficulty with this is that whatever explains why C causes M or why C causes some bodily movement is not going to be in either case the fact that an explanation for something can be given. Further, whatever explains why C causes M should be *of the same general kind* as what explains why C causes some bodily movement. But then what explains why C causes M could not be the fact that such and such explains why C

causes some bodily movement.

Now, in particular, we can note that whatever explains why C causes *M* ought also to explain why C causes *some* bodily movement. But then how can

(*) the fact that C causes some bodily movement because it indicates that *s* is F

explain why C causes M? If it did, it would also explain why C causes some bodily movement. But if (*) is true, then the explanans for C's causing some bodily movement is 'C indicated that *s* is F', and this is not equivalent to (*) itself.

This point is important, so let me try to put it as clearly as possible. Whatever explains why C causes some particular bodily movement M, will *also* explain why C causes some bodily movement. Let `A' stand for `C causes M', let `B' stand for `C causes some bodily movement', let `D' stand for `C indicated that *s* is F', and finally let `E' stand for `C causes some bodily movement because C indicated that *s* is F'. We can now give the following argument:

(1) Whatever explains why A will also explain why B, i.e., if A because X, then B because X.
(2) A because E (from Dretske's analysis).
(3) B because E (by (1)).
(4) E = *B because D*.
(5) B because *B because D* (from (3) and (4).

But (5) is nonsense. For notice that if (5) is true, then its explanans must be true. Thus, D explains why B. But if D explains why B, then B because D does not explain why B, and (5) is false. Thus, if (5) is true, it is false; therefore it is false.

What has gone wrong? The problem here is that two different levels of fact have become confused. There are facts about what causally explains what. And there are the facts about causally relevant features of events and states that do the explaining. The former sort of facts are fixed by the latter. So facts about what causally explains what are not facts about causally relevant features of events and states. Thus, facts about what causally explains what are not

themselves explanatory facts. So the facts *that* will explain why C causes M are not facts *about* what can explain what at all, but facts about causally relevant features of a structuring cause. This is what underlies the difficulty above.

Let us turn now to the question of what does explain why C causes M. On Dretske's own account, it is that C indicated that *s* is F. We need, then, in the place of our explanans above,is `C indicated that *s* is F'. Then we get,

C causes M because C indicated that *s* is F.

Thus, Dretske's analysis makes the representational properties of C the wrong properties to cite in explaining why C causes M.

If we try to correct the problem, however, by saying that

C represents that *s* is F just in case C indicated that *s* is F,

then we must admit that just about everything has representational properties, for anything that has endured for any length of time will have indicated something in the past, and, in fact, many different things. That is why it is so important in Dretske's account that the representation conferring property be that one thing *explained* another, rather than that one thing indicated another, for it is only by this device that he is able to restrict the number of representational systems to something reasonable. But, as we have seen, the very requirement that restricts the number of representational states to a reasonable number at the same time renders representational properties unsuitable for the work of explaining our behavior, however characterized. Thus, Dretske's account is faced with the following dilemma: either the account of representational properties appeals to an explanatory relation in fixing the content of the expression `C represents that *s* is F', and thereby renders it unsuitable for the work of explaining C's causing any bodily movement, or it appeals to properties which are explanatory but shared by virtually everything, so that virtually no state lacks representational properties, which is surely a *reductio* of the position.

4

The moral that should be drawn from the failure of Dretske's account is that we cannot sidestep the threat that neurophysiological explanations will supplant reason explanations. Action explanations are causal explanations. That requires that reasons be causally relevant to what we cite them to explain. Once we see this, we can see that the distinction between structuring and triggering causes cannot remove reasons from competition with neurophysiology, for neurophysiological structuring cause explanations are as available as neurophysiological triggering cause explanations.[6] But the threat that neurophysiological explanations will or should displace reason explanations is exaggerated. There is good reason to think that we should not endorse two or more complete and independent explanations of the same phenomenon.[7] To suppose that reason explanations and neurophysiological explanations could each provide complete and independent explanations of our behavior would be to suppose our behavior is causally overdetermined. We have good reason to suppose that reasons do not operate autonomously from the rest of nature. But the alternative is not accepting that reasons can play no role in explaining our behavior. It is to see reasons as systematically related to neurophysiology, and, in some sense, dependent on it. Reason explanations should "line up" with neurophysiological explanations, and we should be able to explain how they do so in terms of systematic connections between the explanatorily relevant neurophysiological properties and the explanatorily relevant reason properties.[8] Consider the question of what explains why my computer does not sink through my desk. Surely that my desk is a solid object is sufficient to explain this. Yet I could have equally well, if more laboriously, adverted to the interaction of electrostatic forces at the surfaces of my desk and computer. These explanations are not in competition because my desk's solidity is explained by its microstructure, which accounts for the electrostatic forces at its surface. To see reasons as having a role to play in explaining our behavior, we should likewise see them as being explainable by adverting to facts about our neurophysiology. Where further work remains to be done is on what sort of explanatory grounding in neurophysiology would be required to avoid the charge of epiphenomenalism, and whether it is plausible that reason properties could be so grounded.

References

Davidson, Donald. 1963. Actions, Reasons and Causes. Reprinted in Davidson 1980: 3-19.

Davidson, Donald. 1980. *Essays on Actions and Events.* Oxford: Oxford University Press.

Dretske, Fred I. 1981. *Knowledge and the Flow of Information.* Cambridge: The MIT Press.

Dretske, Fred I. 1988. *Explaining Behavior.* Cambridge: The MIT Press.

Dretske, Fred I. 1989. Reasons and Causes. In Tomberlin 1989: 1-16.

Dretske, Fred I. 1990. Reply to Reviewers. *Philosophy and Phenomenological Research.* 50: 819-839.

Dretske, Fred I. 1991. Replies. In McLaughlin 1991: 180-221.

Dretske, Fred I. 1993. Mental Events as Structuring Causes of Behavior. In Heil and Mele 1993: 121-136.

Heil, John and Mele Alfred, eds. 1993. *Mental Causation.* Oxford: Clarendon Press.

Horgan, Terence. 1991. Actions, Reasons and the Explanatory Role of Content. In McLauglin 1991: 73-101.

Kim, Jaegwon. 1989. Mechanism, Purpose, and Explanatory Exclusion. In Tomberlin 1989.

Kim, Jaegwon. 1991. Dretske on How Reasons Explain Behavior. McLaughlin 1991: 52-72.

Kim, Jaegwon. 1990. Explanatory Exclusion and the Problem of Mental Causation. In Villanueva 1990: 36-56.

McLaughlin, Brian. ed. 1991. *Dretske and His Critics.* Oxford: Basil Blackwell.

Stampe, Dennis. 1990. Desires as Reasons—Discussion Notes on Fred Dretske's `Explaining Behavior: Reasons in a World of Causes'. *Philosophy and Phenomenological Research.* 50: 787-793.

Tomberlin, James E. ed. 1989. *Philosophical Perspectives, 3, Philosophy of Mind and Action Theory.* Atascadero: Ridgeview Publishing Company.

Villaneuva, E. ed. 1990. *Information, Semantics and Epistemology*. Oxford: Blackwell.

End Notes

1. In *Explaining Behavior,* most of what Dretske says in characterizing the problem that he is addressing suggests that it is what Terence Horgan has called the problem of the explanatory exclusion of psychology, which is a special version of the a more general problem which Jaegwon Kim has called simply the problem of explanatory exclusion. (See Terence Horgan, 1991, and Jaegwon Kim, 1991.) A later paper of Dretske's (1989) suggests that Dretske's primary concern is with the problem for giving reasons an explanatory role if reasons must be taken to be relational states of an individual. In a footnote to Horgan's (1991) paper, Horgan says that in conversation Dretske admitted that at the time of writing *Explaining Behavior,* he had not clearly distinguished these two problems. In his replies to his critics in (McLaughlin 1991, 203-4 and 210-211) Dretske acknowledges that he did not clearly distinguish between these two problems and agrees that his account should primarily be directed to the problem of giving non-intrinsic meaning or content, meaning or content that is not intrinsic to neurophysiology, a role in explaining behavior. This characterization of his concern, however, does not sit well with the procedure in *Explaining Behavior,* in which a large part of the motivation for giving a relational-historical account of content is that it solves a certain problem about how reasons could explain behavior. This makes sense if the problem is how psychological and neurophysiological explanations could both be true. But it does not make sense as a response to the problem of how to give reasons an explanatory role in explaining behavior in the light of content properties not being intrinsic properties of neurophysiological states. For Dretske offers no reason to think that content properties are relational properties other than the need to make them so in order for them to explain behavior. If the non-intrinsic character of content were the real worry in *Explaining Behavior,* there would not be a problem to begin with.

2. While not so clear in *Explaining Behavior,* Dretske makes this clear in (Dretske 1993).

3. Dretske is not explicit about this in *Explaining Behavior*, but he makes this clear in reply to criticism (Dretske 1990).

4. This may raise protests from philosophers who think that our having the thoughts we do depends on our having had an appropriate history of causal interaction with our environment. The example can be altered to accommodate this, for all that is important is that the person not have acquired his dispositions by having his internal states recruited to cause bodily movements. If five minutes is too short a time, let it be 30 years.

It might still be said that the thought experiment begs the question because it presupposes that we can have representational states in the absence of the kind of history that Dretske's account requires. But the intelligibility of the thought experiment is itself *prima facie* evidence that Dretske's account does not provide necessary conditions for thought content, and this is corroborated by the ease with which we imagine ourselves explaining someone's actions by citing his reasons when we know he has not had any history of the sort that Dretske requires for our explanations to succeed.

5. This was Dretske's response to this objection in a comment on an earlier version of this paper delivered at the 1991 meeting of the Society for Philosophy and Psychology.

6. See (Kim 1990) for a general formulation of the problem for Drestke.

7. See (Kim 1989, 1990).

8. For further discussion of this issue, see my "Causal Relevance and Thought Content," *The Philosophical Quarterly*, 44 (July 1994), in which I argue that the causal relevance of our reasons to the movements of our bodies requires a nomic type-type correlation between the causally relevant physical properties of our bodies, relative to the appropriate circumstance type, and the reason properties we wish to say are causally relevant to those movements.