Philosophy and Phenomenological Research
Vol. L, No. 4, June 1990

Building Belief: Some Queries about Representation, Indication, and Function

STEPHEN P. STICH
Rutgers University

A great virtue of this journal's symposium format is that it gives commentators an opportunity to pose questions to which authors can respond. But there's a down side as well, since the prospect of getting replies "from the horse's mouth" may tempt the symposiasts — it has surely tempted me — to focus on just those points they find most problematic and least clear. So let me begin by setting my questions and critical comments in a broader perspective. Dretske has written a glorious book. It is bold and imaginative, full of insightful and innovative ideas elegantly woven together. It is also a model of how to make important philosophical issues accessible to scholars in other disciplines. The style is lucid, engaging, non-technical and delightfully unpretentious.

The goal of Dretske's book is to provide an account of how intentional states — states like beliefs and desires — can play a role in the explanation of behavior. At the center of that project is what purports to be a thoroughly naturalistic account of the nature of belief. In Section I, I will present a brief sketch of Dretske's story on belief and set out two requirements that his account must meet. Sections II and III will be devoted to raising some questions about a pair of notions that are central to Dretske's analysis.

I.

The guiding metaphor of Dretske's account of belief is borrowed from David Armstrong and F. P. Ramsey: Beliefs are "maps by means of which we steer." (79) A bit less metaphorically, a belief is a representational state whose "semantic character" (*what* it represents) plays a role in shaping behavioral output. (79) To unpack the story further, Dretske has to provide a naturalistic account of representational states, and he has to tell us how their content or semantic character "helps to determine what we

do." (79) The basic ingredients in Dretske's account of representational states are the notions of *indication* and *function*.

Indication is a relation that obtains between a pair of events or states of affairs if the second (the "indicator" or "natural sign") only occurs when the first (the state or event being indicated) occurs. There must be a dependency between the indicat*or* and the indicat*ed* "that we normally express by conditionals in the subjunctive mood." (56) Indicators are everywhere — smoke indicates fire, tree rings indicate growth, and so on, and on. Most indicators are simply ignored. But some end up incorporated into systems in which their *function* is to indicate. A simple example is the bi-metallic strip in a thermostat, whose function is to indicate the temperature. When embedded in the thermostat, the bi-metallic strip not only indicates the temperature, it is also a *representation* of the temperature, since, for Dretske, "a representational system" is "any system whose function it is to indicate how things stand with respect to some other object, condition or magnitude." (52) "What a system *represents* . . . is what these elements have the *function* of indicating. . . . (59) In the case of the thermostat, it is a person, the designer of the thermostat, who determines what the function of the indicator will be. But, Dretske maintains, there are also natural biological systems in which internal indicators have acquired the function of indicating something, typically something about the organism's external environment.

The cases in which indicators acquire their function via conditioning are central to Dretske's story about belief. In conditioning, according to Dretske, an internal indicator, C, comes to be causally connected to another internal state which, in turn, causes the organism to produce a certain sort of behavioral output, M. Moreover, C gets hooked up to M — it gets to be a cause of this sort of behavioral output — *because* M is an appropriate way to behave when the state of affairs that C indicates obtains. Thus when C gets linked to M, it comes to be C's function to indicate that state of affairs. An example will make this a bit clearer. Consider the rat in the Skinner box. When the light is on, pressing the bar will produce a bit of food. Let F be the state of affairs in which the light is on. Presumably there is some internal perceptual state of the rat that indicates that the light is on. Let this state be C. Now, prior to being placed in the Skinner box, C was not causally linked up to bar pressing behavior (= M). But as the result of conditioning, a causal link between the two was forged. C comes to cause M because it indicates F. It is now C's function to indicate F, and as we've seen, for Dretske an indicator of F whose function is to indicate F is a *representation* of F. Moreover, a representation that plays a role in shaping behavior, as C now does, is a belief, or at least a proto-belief.

That, in very brief outline, is Dretske's story about belief. It is ingenious, plausible and disarmingly simple. If the account is to pass muster, then, as Dretske himself emphasizes, it must meet a pair of closely connected requirements. First, the analysis of representation must sustain a plausible account of *mis*representation. Second, the analysis must support a very "fine-grained" or content-specific account of what a given representational state represents.

Now, at the core of my concerns about Dretske's account of representation and belief is the worry that the notions of indication and function will not yield the sort of determinate, fine-grained content that Dretske and the philosophical tradition require. And, of course, to the extent that the notion of representation is not fine-grained or determinate, its capacity to sustain a notion of misrepresentation is threatened. For if it is not clear what a state represents, it may not be clear whether or not it is misrepresenting. (Cf. pp. 69-70)

II.

If Dretske's account of representation is to satisfy the requirements he endorses, then plainly the notion of *function* is going to have a lot of work to do. For, as Dretske notes, the notion of indication is *not* fine-grained, nor does it support a notion of *mis*indication. Moreover, though Dretske does not acknowledge the point explicitly, there is also a sort of indeterminacy built into the notion of indication. Let me take up these points one at a time, starting with the point about *mis*indication.

Indication is reliable correlation — the sort of dependency that is expressed by subjunctive conditionals. If C is an indicator of F, then it must be the case that if F had not occurred, C would not have occurred. Thus if C really does indicate F, then F *must have occurred.* Indication requires that there be some genuine dependency between the indicator and what it indicates, "some condition, lawful or otherwise, that *explains* the persistence of the correlation" between the two. (57) However, as we've already noted, such dependencies are widespread in nature, and it will typically be the case that an indicator stands in an appropriate subjunctively describable relation with many different events or states of affairs. It is, of course, no news to Dretske that indicators are not fine-grained — that they "normally indicate a great many things. . . . " (84, fn. 4) It is the notion of function that is supposed to whittle down the indication relation. While an internal state of an organism will indicate many things, it will represent just one, the one that it is its function to indicate. I'm not at all sure that the notion of function can do the job. But before setting out my concerns on that score, I want to draw attention to an aspect of the notion of indication that Dretske does not clearly

acknowledge. In addition to there being many things that an internal indicator indicates, it also sometimes happens that what an indicator indicates is simply indeterminate; there are cases in which it is just not clear whether or not C indicates F.

Dretske notes that the subjectively describable dependencies underlying the indication relation need not be rooted in universal laws of nature. They may simply reflect the contingent circumstances in the appropriate environment. In Dretske's neighborhood, the doorbell ringing indicates that someone is at the door. "[T]he bell would not be ringing unless someone was at the door." (57) Perhaps in other neighborhoods animals regularly depress doorbuttons. In those neighborhoods, the doorbell ringing does not indicate that someone is at the door. Similarly, "in many cases of biological interest, a sign — some internal indicator on which an animal relies to locate and identify, say, food — will only have this kind of local validity. It will . . . be a reliable indicator only *in* the animal's natural habitat or in conditions that approximate that habitat." (57) However, there will sometimes be an intrinsic indeterminacy about the boundaries of an animal's habitat. And when this is so, the indication relation may also be indeterminate.

An example that Dretske himself uses will serve to illustrate the point. Monarch butterflies store a noxious substance from the plants on which they feed. When a bird eats a monarch butterfly it finds it foul tasting and quickly learns to avoid eating monarchs in the future. Exploiting this fact, another species of butterfly, the viceroy, has evolved to resemble the monarch. However, the viceroy has not evolved the monarch's system of storing a noxious substance. The viceroy is a mimic, an evolutionary freeloader; it gets by on looks alone. Now consider the case of a bird living in a monarch infested environment that is entirely free of viceroys. Presumably, when the bird learns to avoid monarchs there is some internal state, C, that indicates monarchs, and that comes to control butterfly avoidance behavior. The situation is quite different for a bird living in an environment in which both monarchs and viceroys are plentiful. For, assuming that the bird cannot tell monarchs from viceroys by sight, the state, C, which comes to control butterfly avoidance behavior *cannot* indicate monarchs. The presence of viceroys subverts the subjunctive dependency required for monarch indication. C does, however, indicate the presence of *viceroys or monarchs*.

The indeterminacy of the indication relation emerges when we consider cases that are midway between these two. Suppose that a certain species of bird can be found all across the state. At one end of the state both monarchs and viceroys are commonplace. At the other end, there are mon-

archs but no viceroys. And as one proceeds from one end to the other, the density of viceroys decreases. At what point along this continuum does C cease to indicate *monarch or viceroy* and begin to indicate just *monarch*? There is, it seems, no principled way to decide the question. In the intermediate cases, the truth of the relevant subjunctive is itself indeterminate. Cases like this are bound to make trouble for Dretske's account of representation, since if there is no principled way of determining whether C indicates *monarch* or *monarch or viceroy*, then there is no principled way of determining what it represents. When a bird in the middle of the state avoids a very rare viceroy, is it because it misrepresents it as a monarch, or because it correctly represents it as a monarch or viceroy? I do not know how Dretske would answer.

III.

Let me turn, now, to my second worry about Dretske's account: Can the notion of function be relied upon to yield a fine-grained notion of representation? Perhaps the best way to explain why I'm not sure of the answer is to focus on a few examples. Consider first the rat in the Skinner box. Here's how the apparatus works. There is a switch, A, that can be closed by the experimenter. When A is closed, two things happen. A light is illuminated in the box, and a second switch, B, is closed, which connects the bar in the box with the food dispensing mechanism. When B is closed, hitting the bar will deliver a bit of Rat Chow. After the usual sort of training in the Skinner box, a hungry rat will depress the bar as soon as the light goes on. Presumably, there is some internal state of the rat, C, that indicates that the light is on; as the result of conditioning this state has been recruited as one of the causes of the rat's bar pressing behavior. But what does C represent? Perhaps the most intuitively plausible answer is that C represents the light being on. But it is far from clear that Dretske is entitled to this answer. For C indicates many other things as well. It indicates that electric power is being supplied to the lab, and that the power grid supplying the city is in good working order. It indicates that the experimenter is in the lab. It indicates that switch A has been closed, and it indicates that switch B has been closed. It also indicates that it is now the case that pressing the bar will produce a piece of Rat Chow, and that pressing the bar will produce a bit of food. Which of these is it the *function* of C to indicate? Dretske provides us with no analysis of the notion of function. He relies on our intuitive grasp of the concept. But, unfortunately, my intuitions are not up to the task. It seems pretty clear that the first few items on my list can be excluded. It isn't the function of C to indicate that electricity is being supplied to the lab, or that the power grid is in good working order. But what about some of the other alternatives. Why should we not

say that C's function is to indicate that pressing the bar will produce a bit of Rat Chow; or that pressing the bar will produce a bit of food? I have no idea how to go about answering these questions. I hope Dretske will tell us. Until he does, I think we're entitled to be a bit skeptical that the notions of indication and function are up to the task of providing a fine-grained account of representation.

One final point. It might be thought that the case of the rat in the Skinner box is a bit contrived, and that the problem of determining the function of an internal state would not arise in a more natural setting. But that would be a mistake. Consider poor Fido. His master, Clyde, who is usually kind, occasionally comes home drunk. And when he's drunk he has a nasty disposition and regularly beats the poor dog. Fortunately for Fido, when Clyde is drunk he fumbles with the door lock. After some painful conditioning, Fido now hides in the basement when he hears that sound. Let C be an internal state caused by the sound of fumbling with the lock. As in the previous case, C indicates lots of things. It indicates that there is a certain sound in Fido's environment. It also indicates that Clyde is fumbling with the lock, and that Clyde is drunk, and that Clyde is in a nasty mood. In a more pragmatic vein, it indicates that if Fido doesn't hide in the basement, he will get beaten. Which of these is it C's function to indicate? How are we to decide? The intuitive notion of function, or in any event *my* intuitive notion, is not sufficient to settle the matter.[1]

---

[1] Parts of this paper are based on Shawn Lockery and Stephen Stich, "Prospects for Animal Models of Mental Representation," *International Journal of Comparative Psychology*, in press. Thanks are due to Peter Godfrey-Smith, Shawn Lockery and William Ramsey for much helpful discussion.