

# 14. Dretske: A recipe for thought

Martín Abreu Zavaleta

June 17, 2014

## 1 Compasses and recipes

In order to offer a genuine explanation of any given phenomenon, it is not satisfactory to use as the explanandum (the thing to be explained) as part of the explanans (the thing that does the explaining). The point is as obvious as saying that if I want to explain why birds fly, it won't do if I just say that they fly because they have a capacity to fly and sometimes they exert that capacity.

A lot of philosophers think that trying to explain mental intentionality by appealing to things that themselves have intentionality is like the bad explanation above. It's just like offering the explanandum as part of the explanans. However, Dretske makes a case that this need not be so. It would be a bad explanation if we used mental intentionality to explain itself, but it's legitimate to explain mental intentionality in terms of a different kind of intentionality—we may call it natural intentionality.

Dretske makes a case that there is a kind of intentionality that is not exactly like mental intentionality. If this is so, he thinks, it is legitimate to use it as a part of our explanation of mental intentionality:

As long as there is no mystery—at least not the same mystery—about how the parts work as how the whole is supposed to work, it is perfectly acceptable to use intentional ingredients in a recipe for thought, purpose, and intelligence. What we are trying to understand, after all, is not intentionality, per se, but the mind. Thought may be intentional, but that isn't the property we are seeking a recipe to understand. As long as the intentionality we use is not itself mental, then we are as free to use intentionality in our recipe for making a mind [...] (p. 492)

Consider a compass. If it functions properly and is in the right circumstances, it indicates the direction of the north pole. Of course, it doesn't indicate the direction of polar bears, even though the north pole and polar bears are in the same direction. If you take a compass and move polar bears around it, this will have no impact on the pointer in the compass.

An intensional (with an 's') context is one in which coextensional terms can't always be substituted for one another in a way that preserves truth. We often think that intensionality (with an 's') is a good guide to intentionality: if our description of a certain state or condition is intensional, this is good evidence that the state it describes is intentional (it has directedness, as we've put it before). Given that the compass indicates the direction of the north pole but not the direction of the polar bears, and that the direction of the polar bears happens to be the direction of the north pole, it seems that the kind of terms that we use to describe the state of the compass generate intensional contexts. Thus, these descriptions describe intentional states of the compass.

Notice that the intentionality of the states of the compass is not derived from other intentional states. For instance, it's not derived from what we know, what we believe, or our explanatory purposes:

To say that the compass indicates the direction of the arctic pole is to say that the position of the pointer depends on the whereabouts of the pole. This dependency exists whether or not we know it exists, whether or not anyone ever exploits this fact to build and use compasses [...] The power of this instrument to indicate north *to* or *for us* may depend on our taking it to be a reliable indicator [...] but its *being* a reliable indicator does not depend on us. (p. 493)

If this is true, then intentionality is a natural phenomenon, existing in the physical world. The problem is to explain how this kind of intentionality can be used to produce the intentionality characteristic of mental phenomena. In particular, how it is that mental phenomena can manage to *misrepresent*.

## 2 What won't do

Dretske argues that an account of representation appealing only to causation will not do. Consider our explanation of why the compass indicates the north pole. We could say something like this: the compass indicates the north pole because the position of its pointer *causally* depends on the position of the north pole. The pointer moves one way or the other because the north pole's electromagnetic field causes it to move that way.

If we take *indication* or *information* to be defined in such causal terms, we can't explain how something may manage to misrepresent. Any event that is caused by C will thereby be an indicator of C. It just doesn't make much sense to say that something can misrepresent something if representation amounts just to causal dependence. What if there is more than one thing that ordinarily causes *e*, and how should we make sense of an individual event ordinarily being caused by some thing or other?

This gets us closer to another formulation of the problem of misrepresentation, often called the disjunction problem for naturalistic theories of mental representation:

The problem is one of explaining how, in broadly causal terms, a structure in the head, call it R, could represent, say, or mean that something was F even though a great many things other than something's being F are capable of causing R. How can the occurrence of R mean that something is F when something's being F is only one of the things capable of causing R?

Take the compass again. Suppose that there is one very powerful magnet in the same direction as the north pole, and of course, our compass points in that direction. Each of the magnet and the north pole could be causally responsible for the pointer's position, and in this case, it doesn't make a difference to the pointer's position which of the magnet and the north pole are in fact causally responsible, since they are in the same direction. But what, if anything is the compass indicating? Is it indicating the position of the magnet or the position of the north pole?

It's somewhat easy to see how, if we solved this problem, we could solve the problem of misrepresentation. If we had a way of determining which of the two things the compass is indicating, we

would have a way of explaining why, if we move the magnet in a different direction, the compass would now *misrepresent* the direction of the north pole, but would not even attempt to represent the direction of the magnet.

Dretske's point so far is that we can understand the disjunction problem and the problem of misrepresentation as the same problem under different guises. He also makes the further point that, unlike the basic kind of causal indication we saw above, the kind of misrepresentation that the compass can achieve can't be used in our explanation of mental intentionality.

Dretske illustrates with a thermometer and a paper clip. The metal in the thermometer and the clip both carry information about the local temperature: their volume increases or diminishes with the temperature. But it only makes sense of the thermometer to say that it could misrepresent the temperature (if, say, there was something wrong with it). What is the difference?

The only relevant difference between thermometers and paper clips is that we have given the one volume of metal—the mercury in the glass tube—the job of telling us about temperature [...] Since it is the thermometer's job to provide information about temperature, it (we say) misrepresents the temperature when it fails to do its assigned job just as (we say) a book or a map might misrepresent the matters about which they purport to inform us.

But of course, having the job of telling temperature is not something that the thermometer can do by itself. Its function is something that we give to it. Now we are closer to finding what we need in our explanation of thought, or so Dretske thinks.

### 3 Natural functions

Think about our description of the thermometer above. Dretske thinks that if we could find a way of specifying representational functions (or jobs) in a way that doesn't require our intentions, purposes and attitudes, we could combine such functions with natural indicators in order to get the kind of intentionality that characterizes our thoughts. In particular, he thinks that we could solve the problem of misrepresentation. Why? Because something could indicate something that it's not its function to indicate.

He considers two ways in which we could give such a specification, only the second one of which he thinks will help our purposes:

**Evolution:** Some people think that something like the heart has a natural function: to pump blood.

It has that function because of its selectional history: the function of a heart is to pump blood because hearts were selected for their blood-pumping capabilities. Similarly, we may think, our senses have an information-providing function. They were selected for their ability to tell the animal in which they occur about its environment.

**Learning:** Think of an animal that needs to do A in circumstances C in order to survive. Let's suppose that this animal can perceive circumstances C, and that when it does so perceive, it has an internal element E responsible for this. If the animal is to survive, some of the internal mechanisms in the circuit must change so that whenever E is present, the animal performs A. If this happens, then E will have acquired the function of informing of C. Dretske puts the point as follows:

The internal indicators must be harnessed to effector mechanisms so as to coordinate output to the conditions they carry information about [...] internal elements that supply needed information acquire the function of supplying it by being drafted into the control loop because they supply it. They are there, doing what they are doing, *because* they supply this information.

Of these two ways of getting a function, Dretske thinks that only the second one will work. The reason is that the kind of function determined by evolutionary mechanisms can only produce non voluntary responses, but Dretske thinks this is not the kind of mechanism that could explain thought.

Instead, he thinks that the kind of mechanism he dubs *learning* is the one responsible for the right kind of functions. Still, we may wonder whether content defined in terms of this sort of function can escape the disjunction problem. For instance, let's suppose that there is some indicator R that gets its function of indicating cows only by exposure to Jersey cows. This doesn't mean that R means Jersey cow, or just animal, but then how can we make sure that it means just COW? Dretske follows Fodor in claiming that it must satisfy the right counterfactuals. For instance, that if the indicator is presented with a non-Jersey cow, it will still fire or produce the cow-representation.

So what's Dretske's recipe for thought?

Take a system that has a need for the information that F, a system whose survival or well-being depends on its doing A in conditions F. Make sure that this system has a means of detecting (i.e. an internal element that indicates) the presence of condition F. Add a natural process, one capable of conferring on the element that carries information F the function of carrying this piece of information. (p. 497)

Dretske briefly addresses a final objection: suppose that the indicator indeed has a content; in order for it to count as a thought, it must also play the functional role of a thought. In particular, it must be involved in inference and reasoning, and explain behavior. According to Dretske, the indicators with the appropriate natural functions satisfy these requirements: "According to this recipe for thought, nothing can become the thought that F without contributing to a rational response to F, a response that is appropriate given the system's needs or desires." (p. 498)