

Putting Dretske to work

Commentary on Dretske, Fred, "Putting Information to Work"; To appear in Philip P. Hanson (ed.), *Vancouver Studies in Cognitive Science*, Vancouver, Canada: University of British Columbia Press, 1989.

Brian Cantwell Smith

Dretske, I take it, argues as follows. He starts by asking a question:

(1) What is the explanatory role (if any) of information?

The question is made urgent by a problem:

- (2) To be explanatory, something must be *causal*;
- (3) Information, however, is a *semantical* phenomenon, and since (as is widely alleged)
- (4) Causal efficacy is *intrinsic* or *local*; and
- (5) Semantics is *relational* or *distal*;
- (6) It follows that semantics can't be causal (from 4 & 5).
- (7) Therefore, information can't be explanatory. (from 2, 3, & 6).

For example, suppose signal *s* carries the information that *p*, and also engenders behavioural consequence *m*. Then (according to Dretske) *s*'s carrying the information that *p* can't be causally responsible for *m*. Why not? Because they're insufficiently (i.e., not counterfactually) correlated. If you change the content (i.e., make it not true that *p*) without changing *s*, *s* will go on affecting the world in the same way it always did, but will no longer carry the same information (since information is veridical). The two properties – carrying information, and causing effect – are, so to speak, too *disconnected*.¹ And if, as Dretske suggests, this disconnection robs information-carrying of any causal potency, what on earth use can information be? No use, perhaps.

Given question and problem, Dretske then proposes the following solution:

(8) Information derives its explanatory force from situations of *learning*.

The reason? Because during learning, according to Dretske, intentional agents can respond differentially to something's carrying information (or can end up in different states depending on whether something carries information – or some-thing like that). They can end up one way if *s* carries the information that *P*, in other words, and another way if it doesn't. As a result, even if at some later time *s*'s carrying the information that *p* can't be locally responsible for its production of *m*, something else may be true, almost as good: *s*'s carrying the information that *p* can still, at that later date, be causally responsible for the *fact* that *s* causes the production of *m*.

Maybe I've got the types wrong. Maybe facts aren't quite the sort of thing that can be caused. But the intuition isn't difficult to see. If, during the learning situation, *s* hadn't carried the information that *p*, then the agent wouldn't have learned in the way that it did, and therefore (later) *s* wouldn't have caused *m*.

So in this case it looks, according to Dretske, as if semantical properties (like carrying information) can be explanatory after all, because they can explain how symbols get to have the causal powers that they do. Maybe, that is to say, the problem really is solved. And note, too, *how* it would be solved. Dretske proposes to defuse (7) by eliminating (6), rather than (2) or (3). He

¹ Thus, a highway flare's *carrying information about an accident ahead* can't be causally responsible for your putting on the brakes. It can't be responsible because you would have slowed down, upon seeing the flare, whether or not there really were an accident. It's as if the accident itself, though intuitively relevant to why you slow down, is still too far away (in some appropriate sense) to be causally efficacious in making you do so.

claims, that is, that learning is a situation where semantical properties can, after all, be causally efficacious [TPV: effective].

In this review I'll accept Dretske's question. Before considering his solution, however, I want to spend some time examining his formulation of the problem. [Missing from TPV: I've got some questions to ask along the way.]

2. The problem

By my reconstruction, Dretske's problem rests on four assumptions (2-5). Two are relatively unproblematic, but two are going to cause trouble.

Start with those that won't. Assumption (3), that carrying information is a semantical notion, is surely right. People may disagree with Dretske's claim that information is *the* original intentional property, but that it is *an* intentional notion seems hard to doubt.² Similarly, assumption (5), that semantics is relational or distal, can hardly be denied. The spatio-temporal reach of "referring to Sir John A. Macdonald" is beyond the wildest imagination of any known physical force. It doesn't follow, of course, from the fact that *some* semantic relations are remote (all, I take it, are relational), that *all* are remote. The three word term "the name 'Ichabod'", for example, seems to contain its referent right inside it. Nonetheless, the general fact remains: there is more to a system of symbols, representations, or "information-carriers" than can be found internally or intrinsically within them. If you want to determine the truth of the statement in the newspaper claiming that gold's been found in the creek behind your summer home, you don't get out the microscope and peer at the typescript; you go up North and look.³

2.a. The locality of causation

Dretske's other two assumptions, however, are less obvious. Look first at (4) – that causal efficacy arises from intrinsic (non-relational) properties. This may seem tautological, since it's hard not to accept the enduring locality of physical or material interaction. But it isn't necessarily

² In fact the situation is more complicated than this suggests. Dretske himself admits that the real issue – the question we're all finally interested in – has to do with the explanatory force of intentional or semantic notions more generally (including language, representation, meaning, content). He focuses specifically on information not only because he believes it is a relatively clear instance, but more seriously because, by his lights, it is the foundational case (as argued for example in his *Knowledge and the Flow of Information* (Cambridge: M.I.T. Press, 1981)).

This intellectual cartography affects the assessment of Dretske's solution. To start with, the relevance of his specific analysis to the larger intentional question depends not only on specific concerns about its intrinsic viability (of the sort raised in this review), but also on whether information really does play the distinguished role (among intentional phenomena) that he imagines. Furthermore, even if the reader agrees with this much – accepts, that is, that other intentional notions are derivative on information – a separate account of the explanatory value of those other intentional phenomena will still be required in cases where they don't manifest those properties of information on which Dretske's proposed solution rests. Representation is a good example: since it doesn't arise out of the sort of causal dependence that Dretske claims for information (c.f. misrepresentation and representation of non-existent objects), and since Dretske relies on that causal coupling in showing information's utility, it follows that he hasn't even proposed an explanation of how representation can be explanatory.

³ Without a proper theory of intentionality and content, which of course we don't have, it is hard to make these claims precise. Still, the basic idea is simple: "meaning something" or "carrying the information that something" involves not only the sign, symbol, or information carrier, but also the referent, content, or interpretation. This will be denied, of course, by solipsists, nominalists, idealists, "social solipsists" (like Winograd), and any others for whom truth and reference amount to a social form of intersubjective agreement. Since Dretske's metaphysics are at least moderately realist, however, I want to address him from a comparable standpoint.

so. While I'll agree that the locality of immediate, proximal effect is unassailable,⁴ the locality of *causation* isn't nearly as clear, in part because of that notion's recalcitrance. It's famously possible to do something now, like planting a bomb, whose consequences don't happen for a long while, and still to call that long-distance relationship "causal" (if you don't think so, I've got a job for you with an asbestos company). So for discussion I'll use the term "potency" to get at such immediate, local, almost physical properties as can engender discriminable effect, and leave "causation" as potentially more long-distance.

Potency is thus somewhat like "first" or "proximate" cause. But not exactly like. It's different because I mean it to include not only legitimate causes, but also any other putatively non-causal but still immediate influences, such as background or enabling conditions (like gravity). I.e., [TPV: That is,] the aim is to corral the entire set of impinging forces that come together and – mechanically, as it were – give rise to a situation or event. Potency is thus both narrower than causation (due to the locality restriction) and at the same time broader (because of the inclusion of enabling conditions). It's important, furthermore, because something like potency (if not potency itself) is a necessary ingredient in the search for intentional foundations. More sophisticated notions of causation – ones that pack in such notions as relevance, long-distance effect, enabling conditions, triggers, and so on – are ruled out, not simply because of their complexity, but because they are as much in need of theoretic reconstruction as the semantical notions being defined in terms of them.⁵

Focusing on potency is important because it leads to a reformulation of (4):

(4') *Potency* is intrinsic or local.

Rewriting (4) requires rewriting (6):

(6') Semantics can't be *potent*. (from 4' & 5)

This much isn't problematic. In fact (6') seems a better distillation than (6) of the inexorable relatedness of semantic or intentional properties. But now a problem arises. In order to generate the problem (7), we also have to rewrite (2):

(2') To be explanatory, something must be *potent*.

Do we want to do that? Or, rather: is Dretske prepared to accept that revision?

He should, of course, to generate his problem, since it is only the local (potency) version of causation, and its inherent proximity and immediate efficacy, to which the relational reach of semantics stands so pointedly in contrast. Furthermore (and more importantly), if he accepts this reformulation of his problem, he should also honour it in proposing solutions. I.e., [TPV: That is,] it would be unfair for Dretske to propose a solution (i.e., to claim that information is explanatory after all) by relying on a *non-local* form of causation. Similarly, if he discounts as non-causal something that is nonetheless fully and locally potent, on the grounds of its being, say, merely an enabling condition, that will also count against him. Dretske's goal (and we're playing by his rules) is really quite narrow: to demonstrate the local, potent effect of carrying information.

2.b. The identification of explanation and cause

⁴ Ultimately based, I presume, on the locality of physics. It's possible, of course, that the rise of intentionality depends on the sorts of phenomena on which the EPR paradoxes shed such meagre light, but I doubt it.

⁵ My real worry about these more sophisticated notions is that they will turn out to be what one might call "post-intentional": themselves characterisable only in intentional terms (such as the suggestion that "causal" is a predicate on explanations, rather than being a pure ontological or metaphysical category), and therefore barred from playing a role in naturalistic reconstruction.

Dretske's other problematic assumption is (2): the identification of explanation and cause. At issue is the relation between:

- (a) Explanatory relations: between an event or situation a and other events b, such that a figures in the proper, naturalistic, scientific, intellectually satisfying, explanation of why b is the case; and
- (b) Causal relations: between event or situation a and other events b, such that a causes b.

Qua notions – i.e., at the most general level – the two are clearly distinct (the etymology of “because” notwithstanding). Furthermore, to take just one obvious issue, if there are any explanations in pure mathematics, such as why there are only five regular convex solids,⁶ then (at least on a Platonist construal) they must differ extensionally as well, since purely abstract objects presumably don't enter into causal relationships at all. So why are they being equated here? Dretske would presumably defend the move as constitutive of naturalistic reconstruction: whatever the case in mathematics (he would say), the goal is to give a *scientific* account of information, which (he would go on) means a causal account. That identification, furthermore, is made clear in his very first paragraph:

Information isn't much good if it doesn't do anything. If the fact that an event carries information doesn't help explain the event's impact on the rest of the world, then, as far as the rest of the world is concerned, the event may as well not carry information. To put it bluntly, in the way positivists liked to put it, a difference that doesn't make a difference isn't really a difference at all. If an event's carrying information doesn't make a difference – *and by a difference here I mean a causal difference, a difference in the kind of effects it has* – then for all philosophical (not to mention practical) purposes, the event doesn't carry information.⁷

Similar statements permeate the paper:

To put information to work will require understanding how the *causal efficacy* of a signal is altered by the fact that it carries information.

Semantics or meaning, the what-it-is-we-believe (and want) is *causally (and, therefore, explanatorily)* irrelevant to the production of behavior.

How can anyone seriously doubt the *causal efficacy* of information and, hence, its relevance to understanding *why* some things turn out the way they do?

It's obvious that this position is strong, but something else is going to matter as much: the fact that it is a methodological commitment, not an empirical claim. [TPV: This is obviously a very strong position. Furthermore (this is going to matter), since it is a methodological commitment, not an empirical claim, it permeates Dretske's entire analytic stance.]

⁶ Consider a vertex. It must be formed by the intersection of at least 3 planar surfaces (2 would form just a sheet), each a corner of a regular polygon, such that the sum of the angles is less than 360° (360° would make the corner flat; more than that would start to buckle). Start with triangles (60° corners): 3, 4, and 5 are ok (tetrahedron, octahedron, and icosahedron), but 6 makes 360°, and more than that are precluded. Next come squares (90° corners): 3 are again ok (cube), but by 4 you are already at the 360° limit. Similarly, 3 pentagons is ok (dodecahedron), but 4 is too large ($4 \cdot 108^\circ = 432^\circ$). By the time you reach hexagons, 3 (the minimum) is already at the limit (360°), and obviously no higher order figures will fit at all. So that's it: a total of five.

Now was that an *explanation*? Without taking on whole philosophies of science and mathematics, I would say it was. But it certainly didn't traffic in *causation* – at least not any form standing in obvious contrast to semantics.

⁷ The italics, in this and subsequent quotes, are my own.

As for reformulation (2') – that to be explanatorily relevant information must be potent – nothing in the text directly supports it. On the other hand, even these few quotes suggest what his proposed solution makes clear: that the local, immediate, engendering of discriminable effect is exactly what he has in mind as a paradigmatic cause. So I'll call this a vote of acceptance.

2.c. Summary

Taken together, the four (revised) assumptions reveal Dretske's picture of the intentional landscape. One might have thought that local potency, long distance causality, and full-scale explanation lie on something of a continuum – ranging from the immediate or proximal to a realm of wide theoretic compass. On such a view, the salient naturalistic puzzle would be to show how such a continuum of “reach” or “inclusion” could arise. But of course that's not how Dretske sees things. Instead, he presupposes a single, unproblematic, local notion of cause, and then identifies explanation with it. Having thus assumed that all three notions (potency, causation, and explanation) line up together, he asks whether information cannot also be drawn into the same small corner.

3. Dretske's solution

Given this image of his project, let's turn to Dretske's proposed solution (8): that information's explanatory value arises in situations of learning. Just how is this supposed to go?

Structurally, as I said earlier, Dretske's plan is to deny (7), the claim that information isn't explanatory, by eliminating (6), the claim that semantics is causally impotent. I.e., [TPV: That is,] his solution will *be* a solution just in case it demonstrates a case where semantical properties are causally efficacious, after all. That much is clear. But there's still room for some tactical maneuvering. Conclusion (6) can be reversed by denying either of the two assumptions on which it rests: (4), that causation is local, or (5), that semantics is distal. Curiously (the plot thickens! [TPV: the plot thickens]) Dretske's choice isn't clear. Sometimes he seems to opt for one, sometimes for the other.

3.a. The learning situation

The situation we're to imagine involves a rat, an audible tone, and a food dispenser. At an abstract level, we're looking for a case where a signal *s*'s carrying the information that *p* causes *s* to engender behaviour *m*. So the example is parcelled up as follows: *s* is an internal state of the rat's brain, *p* is the sounding of the tone, and *m* is the pressing of a bar. Dretske must therefore show how the fact that the internal rat state (*s*) carries the information (*p*) that the tone is sounding explains why it (*s*, again) subsequently causes the rat to press the bar (*m*).

The intuition goes something like this. Even before the rat has “learned” in this way, it could (by presumption) still hear the tone. Dretske has chosen to register this by saying that the same state *s* occurred (at least in potential) before the learning situation, but did not then cause *m*. During the learning situation, however – through some combination of stumbling around and trial-and-error – things change appropriately, so that when it is all over, *s* *does* cause *m*.

There's nothing problematic with those facts, yet. But then Dretske makes his crucial move. He asserts that the change in *s*'s causal powers comes about because of the semantic relation between *s* and *p*.

In order to understand this claim (I'll get to assessment in a moment), it's important to identify three relevant facts – part of what one might call the causal background (see Figure 1). The first has to do with *s* on its own: what causes it?⁸ The answer is obvious: the tone's sounding – i.e., *p*. Turn on the sound, and *s* will result; turn it off, and *s* will go away. And they are causally coupled

⁸ Or causes it to be activated – it all depends on how you individuate mental states.

in the appropriate way. This is all true, furthermore, both before and after the learning situation. No information needed.

Figure 1: The learning situation [TPV: no such figure appears, just the title of the figure!]

The second question, symmetrically, is about what *s* causes – i.e., with its effects. Before learning, they don't amount to much of anything, we can suppose (maybe the rat simply turns the other way, if the tone is loud or unpleasant). After learning, on the other hand, *s* causes *m* – by hypothesis. Again, there's no need for information; the wiring (control circuits) will by now have been modified appropriately so that *m* simply happens.

The third question is about the effects of *p*. Here the answer is a bit trickier. Since (as I said a moment ago) *p* causes *s*, and (as I said just now) *s* causes *m*, by transitivity of “cause” it would be fair to say that, after learning, *p* causes *m*. It isn't an odd conclusion, either, if you think about it: the tone's sounding causes the bar to be pressed, in virtue of an easily imaginable sequence of appropriately potent states in the rat's head. The causal chain enters and leaves the rat, of course – but there's no reason to feel queasy about that. The rat simply comprises some “middle links” in a connected causal chain. Structurally, the situation is no different from his seatbelt case, where we would (perfectly happily) say that a person's sitting in the seat caused the alarm to go off, explained in terms of a sequence of potent relationships involving switches, wires, and electrical impulses.

So far, things seem rather simple. On the other hand, information hasn't intruded yet. And that's as Dretske plans it. Information, he claims, isn't implicated in causing *s*, nor in causing what *s* causes (i.e., *m*), but in causing *the change in s's causal powers* that takes place during the learning situation, leading it to (thereafter) cause *m*. His words: “the fact that *s* carries the information that *p* ... explains the *recruitment* of internal elements as causes of movement”. And he repeatedly talks of the “redeployment” of *s*'s causal powers. According to Dretske, in other words, information has exactly the following causal role to play: to bring into effect the “higher-order” *adjustment* of the causal powers of a signal that carries it.

So the real thing we are being asked to believe – the putative solution that I want to examine – is the following claim: that whereas *s*, *m*, and *p* are all causally efficacious states, with perfectly ordinary causes and effects (i.e., can on their own be satisfactorily explained without recourse to semantical notions), the *adjustment* in *s*'s causal powers during the learning situation is the *causal* result (assumption 2) of the fact that it carries the information that *p*.⁹

3.b. Causal facts and causal explanations

I see three problems with Dretske's solution. The first has to do, rather directly, with the causal structure of the situation. Think in particular about the learning scenario, complete with its pattern of behavioral modification. And then think about what Dretske doesn't: the food dispenser. Surely what causes the alteration is neither *s* itself, nor any information that *s* carries, but the (causally coupled, not just counter-factually correlated! [TPV: no “!”]) *presence or absence of reward*.

There are at least two reasons to believe this is a better “causal” explanation of the net change than Dretske's. The first has to do with timing. Note that *s* itself can hardly be the cause of the change in its own causal powers – in part because (as Dretske himself says) it has the “same

⁹ Wherever possible I've tried to stick with Dretske's scheme of individuation. . In ordinary usage, the word “because” is ambiguous as between *explanation* and *cause*, although the explanatory reading is probably more common. For Dretske, however, as codified in assumption (2), there's no room for two readings: explanation and cause have been identified. So when he claims that *s*'s change in causal powers arises because *s* carries the information that *p*, he is committed to a causal reading: *s*'s change in causal powers must come about as a causal effect of *s*'s carrying that information. I.e., [TPV: That is,] should you have any tendency to distinguish the questions of *how* the change came about and *why* it came about, you should focus solely on the former. Dretske's claim that *s*'s change in effect is explained by its carrying of information must mean that it was caused by that information carrying.

form” before and after learning. By the same token, however, s’s carrying of information (which Dretske does single out as the relevant cause) is also temporally stable – true before, during, and after learning. If something is unchanged throughout a period, then it alone cannot be called on to explain why something particular happened during the middle.

What is needed, instead, is something (potent) that explains what is *different* about the learning situation. What might that be? The answer is surely obvious: the pattern of activity that includes the offering and retention of food. The rat’s control circuits, speaking very roughly, are “dented” by the (perfectly potent) interaction between its hunger and its pleasurable reaction to food. *Nothing more need be said.*¹⁰

The second reason (why the reward system, rather than the information carrying, is responsible for the change in s’s causal powers) has to do with the potent and quite palpable presence of p. The whole cast of characters, after all – s, m, and p – are all right there, front and center, flexing their efficacious muscles. Dretske sometimes seems blinded by his focus on the intentional fact (that, on his theory of information, s carries the information that p), and thus unable to see what may be much more important: the fact that s and p are not only both causally proximate (to s), but even causally coupled. In fact, as I’ve already pointed out, p is s’s immediate cause. In light of all this potent proximity one is naturally led to wonder whether it isn’t the *causal* relationships among s, p (and the state of the food dispenser) that are doing the work, not the *semantical* relationships among them.

In fact – and this is really the essence of this first problem – you simply don’t need intentional notions at all to explain, locally, what is going on in the situation Dretske imagines. Someone (the experimenter, presumably) causes the rat to be in the situation in the first place, wires up the tone and the food dispenser, arranges for the tone to sound appropriately, and so forth. And then, with the help of a little random exploration on the rat’s part, together with a causally coupled tone and food dispenser, the entire situation can be (causally) explained in terms of the ordinarily efficacious states of bells, ears, air waves, neural circuits, food dispensers, etc. So far, in other words, it’s not just that there’s nothing for information to explain. There isn’t even anything intentional going on.

3.c. Property identity and property overlap

So turn to that: the question of how intentionality enters the picture. This leads to the second problem. It has to do with just what Dretske thinks is the relationship between *carrying information* and *being causally coupled*.

The point is that one must not be misled by the fact that, in this special (learning) situation, *carrying the information that p* and *being causally coupled to p* are simultaneously true. That may be an important fact, but Dretske’s conclusion doesn’t follow from it. The problem stems from a dilemma. Dretske can either claim

- (a) That to carry information is to (have been) causally coupled, or
- (b) That there is more to information than causal coupling.

Unfortunately, (a) is essentially vacuous. Reducing information to causation would automatically satisfy the desideratum of showing how carrying information can do work, but it would do so by evacuating the notion of information of all theoretical interest. *Carries the information that p*, on such a line, would simply be long-hand for *was caused by p* (as would *learning that p*). Premise

¹⁰ By analogy, suppose that instead of “rewarding” the rat by giving it food, the experiment was conducted with an electric rat, and the voltage were simply turned up when the tone was sounded (in fact turned up “by” the tone’s sounding), so that the particular control path whereby the rat pushed the bar was “burned in” more deeply than others, and as a result (imagine some kind of internal contention network) was more likely to be chosen as a result of a series of “training” situations. What we would undoubtedly say caused the alteration in s’s causal consequences, in this situation, over the course of the sequence of trials, would be the effects of the correlated higher voltages.

(5), that is to say, would have been vitiated by the following trick: whereas causing something is local, having been caused by is distal. End of argument. But also end of interest. If intentionality amounts to no more than causation, then the naturalist's task wouldn't be to provide an explanation of information; it would be to eliminate the notion of information from intellectual inquiry.

So what about (b)? It too doesn't work, but for a more interesting reason. Dretske's analysis of information as causally dependent counter-factual supporting correlation rests on a necessary co-occurrence of information-carrying and causal coupling (a co-occurrence that presumably happens early on, if not at the very beginning, of the information carrying period). Quite strikingly, his current account of the explanatory value of information again involves a necessary co-occurrence, this time between information carrying and the complete causal coupling of *s*, *p*, and *m* that obtains during learning. However, in both cases the co-occurrence is (necessarily! [TPV: no '!']) a *sometimes* affair— something I'll call *property overlap*. And property overlap – even necessary (but still occasional! [TPV: no '!']) overlap – isn't property identity. The only thing that would even plausibly give you property identity would be constant co-occurrence. And, to belabour the obvious, it would only be property identity that would justify transferring the higher-order predicate “causally efficacious” from *s*'s *coupling to p* to its *carrying the information that p*. (By analogy, crossing the country on Interstate 80 necessarily overlaps with being in New York City. And being in New York causes a (slight) increase in East Coast smog. But it doesn't follow that crossing the country increases East Coast smog. Being in New York is what does that.)

The importance of property overlap (even necessary property overlap) is that it opens up a middle territory between two properties being *identical* and two properties being *independent*. For example, the size of Maine and the average size of the New England states aren't what would normally be called independent. [TPV: Consider another example: the size of Maine and the average size of a New England state. These two properties aren't what would normally be called independent.] Not only is the average size not independent of the particular, it would be perverse (if not outright false) to claim that Maine's area was independent of the average. On the other hand, it's equally obvious that the two properties aren't identical.

So that's the second problem. It may be true that carrying information and being causally coupled necessarily overlap. It may even be *important* that they overlap (which I believe). Overlap, in fact, may even be *partially constitutive* of learning (though I'm less sure of this). But from none of these facts does it follow, even in the overlapped situations, that *carrying information* does any causal work.

3.d. Long distance causation

Dretske's third problem, separate from subtleties of the learning situation, has to do with exactly what is being claimed. Remarkably enough, he doesn't distinguish the following two readings of his “solution”:

(8a) Information is indeed explanatory, but only during situations of learning [TPV: learning situations].

(8b) Information, since it is causal during learning, is explanatory at other times.

Unfortunately, I think he needs (8b), but at best has argued for (8a).

According to (8a), information does have a use: it lets people learn. Sure enough (the story would go), in ordinary, non-instructive situations – reasoning, say, or action – *s*'s carrying information that *p* wouldn't be able to do any work, and therefore might as well not even be true, which is too bad. Still, at least in a particular kind of instructive setting, information would have a definite, if

¹¹ The situations that Dretske imagines, of course, are far from being *all* learning situations. Learning a skill, learning something on the evening news, learning a person's name – none of these would count as examples of what Dretske has in mind.

limited, role to play.¹¹ This kind of learning, in other words, wouldn't be *what* gives information explanatory force; rather, it would be the only situation *when* it had explanatory force.

I take it that that would be an odd conclusion. First, if it were only during learning that information plays an explanatorily relevant role, then Dretske's own examples – of how strange it is to think that information is useless – would continue to be strange, since those examples don't involve learning. Second, by an argument analogous to the one given above about property overlap, it would be hard (read: impossible) for Dretske to argue that it was *carrying information* that was explanatory, instead of something narrower, such as *carrying information when still causally coupled during learning* (or even just: *being causally coupled*). And third, (8a) claims a modest role for information, but it's not a modest claim; if it *were* what Dretske had in mind, he would have admitted it (which he hasn't).

In fact, however, the text makes it clear that Dretske intends to endorse alternative (8b): that information is explanatory all the time, sustained (in some way) by its causal influence during learning. The intuition is presumably as mentioned in the introduction: that the causal coupling during the learning phase somehow causes the fact that the signal will (later) lead the interpreting agent to respond to it appropriately. I.e., [TPV: That is,] the intuition requires a long-distance notion of causal effect.

But this won't do, for a spate of reasons already cited. In order to avoid evacuating information of substance, there must either (a) be more to the long-distance reach of information than the mere historical record of prior causal coupling, or (b) be more to long-distance causation than the footprint of local potency. But Dretske provides arguments for neither. All his efforts, as I've tried to show, have been dedicated to denying (5), the claim that semantics is causally ineffective. But this direction requires the opposite: eliminating (4), the claim that causality is local. And, as I suggested earlier, this isn't so easy. One would need an account of how explanatory force could stick to information-bearing signals, like pine sap to a wool sweater, long after the learning situation is over – an account, furthermore, that didn't invoke any notions of relevance, enabling conditions, triggers, or any of the other quandaries that plague attempts to develop an adequate theory of genuine causation, and yet amounted to more than the simple persistence of the fact that there was once local causal connection. In addition, if that *were* how the argument was supposed to go, the whole problem would need restatement, since it was only in contrast to a *proximal* notion of causation (potency) that the original intuitions about semantics' relational nature had any bite.

4. Diagnosis

Enough problems. Here's what I think is really going on.

Dretske, I believe, has three things: an insight, a claim, and a method. At least in its intuitive form, I believe the insight is correct (I also believe it is the same one that underlies his original account of information): that there is something crucial about the causal interaction between an intentional structure and its content.

In the present paper he locates this insight in situations of learning. In spite of this, however, so far as I can see, his analysis exploits nothing unique to learning, but would be true of any form of substantial participation in the subject matter, of which learning is merely an obvious case. For example, suppose while driving I maintain a sense of North by checking the dash-mounted compass every so often – more frequently when the roads are particularly twisty. There will be in my head the (maintained) information that *that direction is North*, for some internally oriented sense of "that direction". It overuses the term to call every instance of what is essentially an occasional servo mechanism "learning", but the properties of interest to Dretske still obtain: my internal orientation is differentially dependent on a causally connected chain of events tied to

something that (at least in the arena of interest) is correlated with being North. And the differential result lasts. Perception is involved, of course – but action could be, too. The point is only that some form of causal coupling (that includes causal flow from world to agent) connects the agent to the semantic domain.

Based on his insight, however, Dretske states what I take to be an untenable claim: that (in consequence of this causal coupling) *semantical* relationships are endowed with *causal* powers. For all the reasons cited, I don't think this follows. In fact there seems to be a danger that Dretske has the project backwards. He tries to show that an intentional relationship must exert causal force, whereas all that his examples demonstrate is that causal couplings are necessary ingredients in establishing intentionality.

The mistake, furthermore, is based on a common confusion. Dretske starts with the pervasive intuition that semantical relations aren't causally efficacious. And he concludes – as many people conclude – that this implies that thought, reasoning, computation, and similar intentional phenomena must proceed *independently* of their semantic value or content. This conclusion is reflected in the universal tendency to say (as Dretske does) that “you can change the semantic value, without changing the symbol”.

This doesn't follow. If, as I've suggested, intentionality is inherently participatory (learning is just one example), then you can't necessarily perform that change. If some causal properties are even partially constitutive of intentionality, that is to say, then to change the semantic value may require changing those properties. This was the lesson of overlap. For example, the symbol might have to be propped up appropriately. Imagine a meter wired to a radar telescope indicating the exact current distance to the moon. We all know that this meter can fluctuate and flap around without being connected to the moon – as it was while being designed and tested, for example. But to make it indicate other things (as its designers and testers did during construction), you have to *wire it up to something*, since that's how it's got to be driven.

I.e., [TPV: That is,] there are cases – more or less, I haven't yet said – where various kinds of causal connection play an essential (but not total) role in engendering intentionality. But can we conclude from that that semantical relations are causally potent? *No!* Only that causality is – sometimes, and maybe even then only partially, but still, to some extent – constitutive of intentionality. I.e., [TPV: That is,] you can challenge the independence of semantical and causal relations, without thereby committing yourself to thinking that intentional properties are themselves causal – just as I am not my arm, though I am not independent of my arm, either.

The result, or claim, that causal coupling is partially constitutive of intentionality, is important, I believe. Terrifically important. It is legitimate to ask, however, whether it is *new*. Causal theories of reference, for example, are presumably founded on something like the same intuition. And so is the near-universal assumption (even Fodor believes this) that semantics is grounded in perception, action, etc. What would be new would be a satisfying theory showing just what it came to, how it all worked. But none of this requires that we do what Dretske did: *reduce* information to causation. Nothing about how causation is constitutive of intentionality, that is to say, requires that we give up on what I take to be the deepest intentional insight of all:

Genuine semantical relations – including information – outstrip the locality of (whatever it is that we think we get at with the word) “causation”.

But does this maxim imply that information isn't explanatory, after all? *No, not that either!* And this brings us straight back to what I think is the deepest problem with Dretske's whole argument. The problem is with assumption (2): that explanation can reduce to cause.

Here, finally, we get to Dretske's method – and to the *reason* why he is forced to a causal reduction of intentionality.

If you restrict explanatory relations to causal relations, it's not just that Dretske has failed to show that information can “do work” – no one else could, either. I.e., [TPV: That is,] in the sense in

which Dretske sets things up, information *can't* have any explanatory value. *But it isn't right to restrict things in this way.* In fact to do so, I would argue, is not just reductionist, but fatally so. Fatal in that it doesn't make room for the recognition that local, causal couplings can, sometimes, lead to a system's overall exhibition of properties that *can't be locally identified.*

Various writers, worried about the relation between one level of explanation and another (often, between intentional and physical levels of description), have distinguished between *reduction* and *supervenience*, where in the first case the predicates and terms of the "higher-level" theory are directly translatable into the predicates and terms of the lower-level one, whereas in the latter they are not, even though the phenomena described by the higher-level account in some sense still "rise up out of" those described at the lower level. What Dretske has implicitly but nonetheless strongly done, by setting up his project in the way that he has, is to bypass the freedom offered by the supervenient line, and *endorse a full theoretical reduction of the intentional in advance.* Dretske, that is to say, is methodologically committed to reductionism.

Which I don't believe is right. For example, suppose a student asked you why two metronomes, mounted in the same (relatively rigid) base, and both obviously vibrating, weren't beating against each other, but were instead producing a single pure tone. By way of answer, it would be perfectly legitimate to say that they did so because they were synchronised. All that naturalism requires is that *you be able to tell a causal story that showed how that synchrony arose.* It does not require something stronger (and something impossible): that synchronisation can count in an explanation of the single tone *only if it can be shown to be a causally efficacious property* (which I assume it isn't).

So what are the semantical morals? These three:

- (1) Learning is important, but participation, of which it is a species, is what really matters.
- (2) Causal coupling (and Dretske) notwithstanding, semantical reach still outstrips causal reach.
- (3) Naturalistic explanation may rest on causal potency, but it doesn't reduce to it in anything like as direct a way as Dretske imagines.

In sum, then, Dretske [TPV: Dretske, in other words,] first reduces explanation to causation, and then tries (unsuccessfully) to do the same to information. I think he can't do the latter, and shouldn't do the former. So what should he do, instead? Well, here's what I take to be our collective homework. First, we need a naturalistically satisfying explanation of how intentionality manages to outstrip causation. With that in hand, we'll be ready to show how explanation can do so too.*

* Thanks to Kathleen Akins and David Chapman for comments on an earlier version.

[Between the text of footnotes 9 and 10: *With respect to the question on the table, there's no room for ambiguity. What there is room to question is Dretske's answer.*]

Reply to Dretske's "Putting Information to Work" Putting Dretske to Work Page Printed at on
Copyright © 1989 Brian Cantwell Smith Page Draft! Please don't distribute, cite, or É Draft!
Copyright ©1988 Brian Cantwell Smith Last edited: February 20, 1989 – 22:55:16