

This work originally appeared in:

Dennett, Daniel. 1990. Ways of Establishing Harmony. In *Dretske and His Critics*, ed. B. McLaughlin, 118-130. Oxford: Blackwell Publishing.

It is not available electronically from the publisher.

This is Daniel C. Dennett's final draft before publication. It has been modified to reflect the pagination of the published version of the work.

Ways of Establishing Harmony

By Daniel C. Dennett

Folklore has many tales in which the hero goes in search of some precious thing - a holy grail, a magic talisman - and eventually, thanks to his many virtues, acquires the sought-for treasure but only after rejecting it, unrecognized, several times. In the final denouement, he is shown that what he wanted so badly he had already had within reach.

One can see in Fred Dretske's career just such a quest - with just such a happy result. I am pleased to be the hearer of glad tidings, especially since, in Dretske's particular saga, he has often mistaken me for one of the villains (or better: worthy opponents), rather than the protector, all along, of the magic key. What Dretske has sought, with increasing clarity of purpose over the years, is the key to how meaning makes a difference to us, and now that he has described his quest so vividly, the rest of us can see better what we were trying to say and do as well.

In a recent paper (1985, p. 31), Dretske gives a particularly clear and telling example of the difference that he is trying to capture: the marijuana-detecting dog whose tail wags because an event occurs in its brain "meaning" (carrying the information) that marijuana is present, but which doesn't wag because the event means what it means. Unlike the dog's tell-tale tail-wag, Dretske insists, our bodily actions often happen because of what the states that cause them mean: "it is the structure's having this meaning (its semantics), not just the structure that has this meaning (the syntax), which is relevant to explaining behavior" (personal correspondence, quoted in Dennett, 1987, p. 307).

In a more recent paper (Dretske, SOFIA conference, 1988), on which this will focus,¹ he puts it even better: he has been "increasingly pre-occupied with the question, not of what meaning (or content) is, but what meaning (or content) does." He wants to give meaning "some explanatory bite," (p. 5 of Dretske's conference paper, hereafter MS) and ideally, this would involve showing "the way meaning can figure in the explanation of why" an event causes what it does.

What difficulty lies in the way of this goal? The arguments, from myself and others, to the effect that the meaning of a particular event always must be, in Dretske's terms, epiphenomenal with regard to that event's

actual causal powers. In *Content and Consciousness* (1969) I argued that the intentional interpretation of neural events is always at best a "heuristic overlay", and in "Intentional Systems" (1971) I tried to show how physical-stance predictions of the effects of structures always had hegemony over predictions based on the powers attributable to those structures (in idealization) in virtue of their meaning - as discerned from the intentional stance. More recently (1981, 1983, 1987), I have spoken, as Dretske notes, of the impotence of meaning; the brain is first and foremost a syntactic engine, which can be fruitfully viewed as reliably mimicking a semantic engine, but in which meanings themselves never overrule, overpower, or so much as influence the brute mechanistic or syntactic flow of local causation in the nervous system. (A semantic engine, claim, is a mechanistic impossibility - like a perpetual motion machine, but a useful idealization in setting the specs for actual mechanisms.) Others have made similar claims: Fodor has long insisted on the inescapability of what he calls the Formality Constraint, and Dretske attributes to Schiffer (1987) the view that meaning is an "excrescence" which can do nothing and explain nothing.

Dretske's response to these claims is ambivalent. On the one hand he makes it entirely clear that he would truly love to defend a doctrine of brains as real semantic engines, with real meanings locally throwing their weight around and making the most direct imaginable difference - but he knows better. He is convinced by the arguments that show, in LePore and Loewer's terms, that the historical facts on which the meaning of a structure supervenes are screened off from the explanation of the structure's causal powers and behavior. (In a similar vein, I have spoken of "inert historical facts.") In fact, his own presentation of the argument, drawing the analogy to the difference between value and perceived value, is the clearest yet. He has reached the First Camp, and he finds it a hateful place but one he must endure now that he has cleaned it up.

What he will not endure, however, is any of the paths to the Second Camp. The well-known responses to the Lesson of the First Camp all involve finding some more indirect way in which there can be a regular, reliable correspondence between the meaning of a (neural) structure and its effects on behavior. For instance, Fodor's language of thought hypothesis is essentially an attempt to describe a system that satisfies what Haugeland has called the Formalists' Motto:

If you take care of the syntax, the semantics will take care of itself.
(1985, p. 106.)

This motto lies at the heart of what Haugeland calls GOFAI (Good Old Fashioned AI) and I call "High Church Computationalism" (Dennett, 1986). There are other, more noncommittal, ways in which materialists who have reached the First Camp can postulate a reliable correspondence between semantics and mechanism, but Dretske finds them all intolerable. I think it is wonderful that he chose to illuminate this "corner

materialists have painted themselves into" by some old-fashioned labels: "epiphenomenalism" and "pre-established harmony." These dreaded relics of prescientific philosophy of mind, these desperate and doomed escape routes from Cartesian dualism, are sure to strike terror in the heart of complacent materialists, and Fred tells us exactly why he cannot abide them: any view of meaning according to which there is merely a pre-established harmony between the causal facts and the facts of meaning may permit us to predict an agent's behavior, and control an agent's behavior, but will not permit us to explain an agent's behavior - "and that," he says quite reasonably, "is what it takes to vindicate belief-desire psychology or our ordinary view about the causal efficacy of thought - that we stopped, for example, because we thought the light was red."²

But since Dretske has reluctantly accepted the conclusion of the First Camp, he realizes that a straightforward vindication of this intuition is not in the cards. What he offers us instead is an attempt to salvage, if not locally potent meanings, then the next best thing: "the fact that A means M, though it fails to explain why B occurred, may help explain a closely related fact, the fact that events of type A, when they occur, cause events of type B. And this fact, especially when we are trying to explain the behavior of a system, is a fact eminently worth explaining" (p. 11 of Dretske's conference paper).

What we need, in short, is not just a brute pre-established harmony, but an explanation of why and how the harmony is pre-established. Moreover (if Dretske has his druthers) this explanation will make an ineliminable appeal to the meanings of the elements thus linked. Now apparently he thinks that we pre-established harmony guys have failed to offer such an explanation, for he is forthright in his opposition: "I don't think this works. Or, if it does work, it does so at a cost that I'm not prepared (unless forced) to pay" (p. 9 of Dretske's conference paper). But this is where he makes his mistake, for in the end, he does not offer us an alternative to pre-established harmony, but a version of it, a truncated version, in fact, of the very version I have offered. (We have reached the turning point in the saga. Reach out, Fred, and take the key!)

There are exactly five ways in which such a correspondence a "pre-established harmony" between the meanings of structures and their causal powers - could (in principle) come into existence. Fred encounters them all, but fails to recognize them for what they are.

First, there are the Three Cousins:

1. the correspondence is designed by natural selection;
2. the correspondence is designed by a learning process of some sort in the individual brain;
3. the correspondence is designed by an engineer creating an artifact, such as a robot or computer.

Then there is the Philosopher's Fantasy:

the correspondence is the result of a Cosmic Coincidence

Finally, there is the Theologian's Hope:

the correspondence is created and maintained by God.

Desperate though he is to wed meaning and causation together. Fred rightly dismisses this fifth possibility with only a passing allusion, for the obvious reason that it would be, quite literally, a *deus ex machina*. It is interesting to me that philosophers who would be embarrassed to spend more than a passing moment dismissing this fifth alternative are nevertheless irresistibly drawn to extended discussions of the implications of the fourth, Cosmic Coincidence, which is actually a more fantastic and negligible "possibility in principle."

Notice that there really cannot be a sixth route to pre-established harmony. If such a harmony is not just a single, large ("Cosmic") coincidence (4), or a miracle (5), it must be the product, somehow, of lots of tiny, well-exploited coincidences, for the moral of the First Camp is that meanings cannot directly cause things to happen - and hence they cannot directly cause themselves to correspond to any causal regularities in the world. So it will have to be via an indirect process of fortuitous coincidences that are duly "appreciated" or "recognized" or "valued" or "selected" by something - either something blind and mechanical, such as natural selection or operant conditioning or "neural Darwinism" (Edelman, 1988) (1 and 2), or something foresightful and intelligent, such as an engineer (3). Any such process is a design process, and must consist, at bottom, of such generate-and-test cycles, where the generation of diversity to be tested is somewhat random or coincidental (see "Why the Law of Effect Will Not Go Away", Dennett, 1974, 1978).

Let me preview Dretske's encounters with the Ways, before setting out the details: he endorses the first two Cousins, natural selection and conditioning, while rejecting the third Cousin, engineering, never recognizing that by his own principles they are inseparable kin who stand or fall together. He then uses his allies, the first two Cousins, to defeat the fourth Way, Cosmic Coincidence, which really doesn't pose a threat in the first place. What clouds our hero's vision? As we shall see, it is his overzealous vow of allegiance to the intuitions of folk-psychology.

Now the details. Dretske begins his attempt to give meaning explanatory bite by nicely distinguishing between explaining B's happening (A caused it) and explaining A's causing B. He then tells a story about how a neural assembly can come to have a meaning, either via species evolution or via phenotypic learning - a kind of intra-cerebral evolutionary process of conditioning. The upshot is that such a story can explain why it is that As, meaning what they do, cause Bs. So far as I can see, this account follows precisely the path I laid out in *Content and Consciousness*, in my discussion of "the evolution of appropriate structures" (pp. 47-63).

Dretske describes an organism with a need to develop an avoidance mechanism against the highly toxic condition F, while I described different strains of organism which were different wired up to avoid, approach, or ignore a particular stimulus condition that happens to be "more often than not" - an important proviso - injurious (p. 49). The result of both thought experiments is the same: the lucky ones who happen to be wired up to avoid the toxic, *ceteris paribus*, are the ones who survive to reproduce, their coincidentally happy correspondence being selected for by natural selection. Dretske goes on to suggest that the same result could also be achieved by redesign during an individual organism's lifetime via a conditioning process, or as I put it, via a process of "intra-cerebral evolution" which can, in principle, yield all and only the new designs which the much slower process of species evolution could produce (pp. 56-63).

Note, however, that while Dretske now agrees with me about this, his doing so requires that he abandon, or at least soften, the hard line he has previously taken on these issues, in *Knowledge and the Flow of Information* (1981), and more recently in "Machines and the Mental" (1985) and "Misrepresentation" (1986). In his book, he attempted to erect meaning on a foundation of information. That is, he developed an account of semantic information (the sort of meaning needed for psychology - a functional notion that applies only to designed channels) from a base of nonfunctional, nonsemantic information channels, through which traveled items with "natural meaning" - items that informed (by definition) with perfect reliability; where natural meanings are concerned, no misrepresentation is possible. The task then was to turn the corner, somehow, from natural meaning to semantic or "natural functional meaning" (in which misrepresentation is possible), without, as he has more recently put it, "artificially inflating" the attributions of meaning to a structure one is so interpreting. Dretske has tried to hold the line against inflation by insisting on what he now calls the "indicator relation." But, although the current paper is not forthright about it, the indicator relations he now endorses (thanks to his endorsement of natural selection as a Way) can only approximately carry information about the distal conditions one is tempted to say they are designed to inform the organism about.

The indicator relation, which he heralds as "a plausible, at least a possible, partial basis for meaning" need only be a rough-and-ready guide to the meaning of the chosen structure. At least Dretske ought to recognize this, for that is how evolution, at the species or neural level, works. Mother Nature is a stingy, opportunistic engineer who takes advantage of rough correspondences whenever they are good enough for the organism's purposes, given its budget. Fred correctly wants to boast that his account does find an explanatory role for meanings: it is because of what a structure happens to indicate that it is "selected for" or "reinforced in" a further causal role in the economy of control of the organism, but it will be selected for, or reinforced in, this role whenever it is "close enough for government work" as the engineers say.

The most Dretske can get out of Cousins (1) and (2), in, is an explanatory account that makes ineliminable appeal to quasi-indicator relations or approximate meanings. Mother Nature never holds out for high-priced indicator relations.

As my frequent allusions to engineering suggest, and as I have argued in *The Intentional Stance* (chapter 8), the design processes one encounters in (1) and (2) are not only fundamentally the same, but fundamentally the same as the artificial design process encountered in (3). It is striking that Dretske resists this conclusion, most emphatically in "Machines and the Mental" but still strongly in evidence, with somewhat difference emphasis, in the current paper. If it is an engineer or computer scientist rather than natural selection or a learning history - who does the selecting (who esteems a structure for its quasi-indicator relations and harnesses it in a particular functional role in a robot's control structure), this somehow gives the structure an illegitimate ancestry, in Dretske's eyes. But why?

Why should selection by an engineer disqualify a structure for one of Dretske's why-explanations, while "natural" selection does not? One might even suppose that engineer-selection had an advantage over its "blind," inefficient competition. One might say: whereas one must use scare-quotes when talking of natural selection's "appreciation" of the potential meaning of a structure, engineers sometimes really and truly respond to these potential meanings in the course of their conscious, deliberate, designing. But Dretske does not find this line of thought palatable, in spite of the fact that in the present paper his illustration of the wiring of the light switch explicitly appeals to just such different appreciations or intentions on the part of circuit designers. Is it just because he has a bad case of orqanophilia (otherwise known as silicophobia)? I think not. I suspect that he is distracted by an illusion: the illusion that somehow in ways (1) - and (2) - but in (3) not - the organism itself (or even: its mind or soul) does the understanding - responds directly to the meaning. After all, one might naturally but confusedly say, if some engineer is responsible for appreciating the meaning of a neural or computer structure, and is responsible for designing its further role in the light of that meaning, there is nothing left for the organism or robot to do - no task of understanding left over. But of course if that were so, the very same consideration would disqualify natural selection, for if it is the process of natural selection that has set up an innate correspondence between meaning and causation, there would likewise be no supervisory role for the organism (or its central, understanding homunculus) to play. (For more on this see my commentary on Ewert's "Prey-Catching in Toads," in *Behavioral and Brain Sciences*, (1988) "Eliminate the Middletoad!")

Now the reason I suspect that Dretske is beguiled by this line of thought is that in earlier work, he endorsed only (2), a learning history, and resisted endorsing (1), natural selection, as a legitimate Way to Harmony.³ Both in his book, and more recently in his articles (and in personal correspondence, quoted in *The Intentional Stance*, p. 306),

he held the line against innate, unlearned meanings. "Beliefs and desires, reasons in general (the sort of thing covered by the intentional stance), are (or so I would like to argue) invoked to explain patterns of behavior that are acquired during the life history of the organism exhibiting the behavior (i.e., learned)."

Why then has Dretske been tempted to favor individual learning histories exclusively? Because, apparently, they seemed to give the organism itself a proper role in the acquisition and appreciation of the meanings in question. But once again, of course, if "mere conditioning" is responsible for the redesign of the individual organism's brain, this too looks like taking responsibility away from the inner understander that Dretske is so loath to lose.

"But where does the understanding happen?" One might ask this, having been treated to an account of the way in which an organism was caused, thanks to its design, to respond appropriately (first internally, and eventually externally) to events impinging on it. It is as if one held out the hope of locating a moment of understanding by using a more powerful microscope, but in fact the understanding is, must be, located more diffusely, in the happy interactions of all those well-designed mechanisms.

This is, I think, the primary source of Dretske's earlier (and continuing!) opposition to natural selection and his continuing opposition to engineering design as sources of harmony. In fact, in recent correspondence with me, he made it quite clear that he put so much stock in learning history, that he was prepared to grant real meaning even to the structures in an artifact, so long as they were produced, in part, by individual learning by the artifact: "I think we could (logically) create an artifact that acquired original intentionality, but not one that (at the moment of creation as it were) had it" (personal correspondence, quoted in *The Intentional Stance*, p. 305).

Now here we must distinguish two claims, one plausible and important, and the other - Dretske's - obscurely motivated and, I think, strictly negligible. The plausible and important claim is that it is astronomically unfeasible to create, by the usual engineer's methods, the sorts of structures that are naturally and efficiently created by learning histories of the sort he champions. This suspicion strikes at the heart of a particular fantastic hope of some in AI, who would hand-craft the myriads of beliefs that would constitute the "world knowledge" of an adroit robot. Not only have some thought this was possible in principle; there is a multi-million dollar project in AI with that as its explicit goal: Douglas Lenat's CYC project, an effort which Lenat himself supposes will take person-centuries of programming to accomplish. (See my discussion in *Daedalus* (1988a).) The majority opinion in AI, however, is that this is a hopeless approach, for reasons well-canvassed by David Waltz (1988).

Dretske's point, in contrast, is philosophical, not practical. It is that even if the engineers could hand-craft all those structures, they wouldn't have any meaning until they had been somehow annealed in the fire of experience. He puts the astronomically unfeasible product of engineering design

in the same category with the even more astronomically unlikely case of Cosmic Coincidence, in spite of the fact that in the former case, there would be explanations of the provenance of structures that made appeal to meanings. He imagines that a physical duplicate of himself might "materialize - miraculously or randomly - out of some stray collection of molecules" (ms p. 19), and goes to some length to insist that this biological twin's motions would not be actions, with meanings behind them.

I move my arm in this way in order to frighten away a pesky fly. With such a purpose I am, let us say, shooing away a fly. That is my action. My biological twin, though he moves his arm in the same way (with the same result) does not shoo away a fly. He doesn't have wants or beliefs, the kind of purposes I have in moving my arm. He isn't therefore, performing the same action. (ms, p. 19)

Your intuitions may agree with his, or recoil, but in either case, they concern something negligible, for as he goes on to acknowledge, there is a loophole: this metaphysically shocking state of affairs is apt to be short-lived, since it will persist only "until the twin accumulates enough experience - until, that is, his internal processes acquire the requisite extrinsic relations - to give his control processes, the processes governing the movement of his hand, the same kind of explanation as mine" (ms, p. 20).

How long, one wonders, should "acquiring the requisite extrinsic relations" take? I should think it would be instantaneous.⁴ Signals from the bio-double's peripheral vision (or perhaps a faint blip from cutaneous sensors on the shoulder) happen to put the bio-double into what would be a bogus fly-out-there-sensing state - except that this time it is caused by a real fly. Moreover, the real fly's trajectory intimately determines the hand-eye coordination series that promptly leads to the (bogus or real?) "shooing" motions. How many flies must buzz around the head of a bio-double before he can start shooing them? If that isn't an angels-dancing-on-the-head-of-a-pin question, what would be?

This curious question, of how much traffic with the world is enough, somehow, to ensure that genuine meaning has been established, is simply the enlargement (via a Cosmic Coincidence) of the curious question that has bedeviled some evolutionary theorists: how much selection is required to endorse a tiny coincidence (a random mutation) as a genuine adaptation? (See my discussion in *The Intentional Stance*, pp. 320-1 and footnotes on those pages, and in my response to commentary in *BBS*). But if nothing but arbitrary answers (e.g., 42 generations of selection) could "settle" the question for natural selection, only arbitrary answers (e.g., 42 flies must buzz) could settle the question for a learning history, for the processes have the same structure they must begin with a fortuitous or coincidental coupling, thereupon favored -- and they have the same power to design structures in indirect response to meaning.

One might hold out some hope of a less arbitrary answer in the special

case of design by conscious, deliberater engineers, for one could require that a structure could acquire only those meanings that were "explicitly recognized" by the designers, but such a reliance on the Word would be of dubious value since, as I argued in *The Intentional Stance* (pp. 284-6), and as Dretske has also shown in his current arguments in support of screened off history, the results of the three different processes can be indistinguishable: there is no surefire way of telling, for instance, that some correspondence in your own brain is not the result of ancient tampering by visiting extraterrestrial bio-engineers.

There remains a marginal sense in which (1) and (2) are more "natural" ways of establishing meanings, but this is a sense of no theoretical importance. In just the same sense, a weasel's short legs are natural, while a dachshund's, as a product of artificial selection, are not. It is an interesting historical fact, but it makes no difference to the integrity of any explanation that appeals to the function, or meaning, or implications, of those features.

This should leave us, and Dretske, with three equally legitimate Ways to Harmony, and indeed he comes within a hair's breadth of granting this. Consider the fundamental claim that underlies his paper: his analysis of why the meaning of a structure cannot figure in causal explanations of its effects:

If meaning supervenes, at least in part, on the extrinsic properties of an event - historical and relational facts that need not be mirrored in the event's current (= the time at which it has its effects) physical constitution or structure - then if A causes B, then the fact, if it is a fact, that A means M will not - indeed, cannot - figure in a causal explanation of B. It cannot because, in similar circumstances [*my italics*] an event lacking this meaning, but otherwise the same, will have exactly the same effects. So it isn't A's having meaning M that explains why B occurred. (Ms p. 10)

This would be a significant worry if "similar circumstances" were apt to be forthcoming with non-zero probability in cases where A did not mean M, but we can rest assured that we will never encounter such an anomaly. That is to say, it is no accident that events with the meanings they have get to play the causal roles they play (and Fred gives a good account of this), but the other side of that coin is that the odds are astronomical against the occurrence of an event or structure that lacked the relevant meaning somehow arising to cause a bogus B-type event. We don't need any special history of experience to ensure that any harmonies we encounter are robust, and Dretske's own arguments show that a learning history, "acquisition of extrinsic relations," could not impart any intrinsic power to those harmonies.

Let me review Dretske's view, and compare it with the view he says he will not accept, unless forced to. On Dretske's view, an account of the historical process of "acquiring the requisite extrinsic relations"

explains why structure types (As) cause their associated Bs. Enis is fine, even if he slights some perfectly good varieties of historical process. But, to quote Dretske's earlier remark: "Meaning, on this view of things, has to do with the etiology of a structure type, with how (and perhaps why) it was developed to service an organism's needs." (ms, p. 7). Dretske ends up showing how, even if meaning cannot be locally potent, it can be put into reliable pre-established harmony with locally potent structures.

So I think Dretske and I end up, if not yet in perfect harmony, at least in possession of all the same themes with which to explain how and to what extent - meanings can be cited in the explanation of events.

Postscript

So matters seemed to me until Dretske's Postscript arrived, recanting his endorsement of natural selection (or more fairly, denying that he had ever meant to endorse natural selection as a Way to Harmony). He gives this account of the difference between a learning history and a selection history:

Natural selection gives us something quite different: reflex, instinct, tropisms, fixed-action-patterns, and other forms of involuntary behavior--behavior that is (typically) not explained in terms of the actor's beliefs and desires (if any). These genetically determined patterns of behavior often involve (as triggers for response) internal indicators (information-carrying elements), but, unlike, belief, it isn't their content that explains the way they affect output. That is determined by the genes.

What exactly is the contrast?

In order to get meaning itself (and not just the structures that have meaning) to play an important role in the explanation of an individual's behavior (as beliefs and desires do) one has to look at the meaning that was instrumental in shaping the behavior that is being explained. This occurs only during individual learning. Only then is the meaning of the structure type (the fact that it indicates so-and-so about the animal's surroundings) responsible for its recruitment as a control element in the production of appropriate action.

The only difference I can discern, however, is that the "structure type" in the case of natural selection is a type that is identified in the genotype, while the structure type in the case of intra-cerebral evolution is a type that is identified only in the phenotype. In both cases "meaning was instrumental in shaping the behavior" - that is, in shaping the behavior-type, and in neither case was meaning instrumental in shaping any particular, individual token of a behavioral type.

Let us compare the particular case Dretske describes his shoeing of the fly -- with a near neighbor produced, presumably, by an evolutionary history rather than a learning history - the shoeing of a fly by a horse swishing its tail. Whatever the actual facts about such cases, let us stipulate that Dretske's arm-swing is a genuine action, for the reasons he gives, and that the horse's superficially similar tail-swing is a mere reflex, genetically programmed.

The horse's tail-swing belongs to a type which is designed to be provoked or caused by various other internal types, normally produced by flies buzzing around its rump; if these mediating types hadn't meant what they have meant over evolutionary history, the horse wouldn't be wired up to be caused by them now to issue one of the tail-swing type behaviors. But of course, right now, given the way nature has set up the horse's wiring, a particular event of one of these mediating types however it in particular is caused, and regardless of what, if anything, it in particular means, will cause a tail-swing. That's the way it is with reflexes: no locally potent meaning.

Is there a contrast, however, in the case of Dretske's shoeing action? In virtue of his individual learning history, he has come to acquire various beliefs and desires, among them the belief that when he sees or feels something roughly like that, it is apt to be a fly or bug or something in need of shoeing, and a desire to shoo whatever needs shoeing. Such belief-desire pairs, when provoked by particular circumstances, are designed to cause the formation of shoeing; intentions which forthwith get executed. (The particulars of this rough-and-ready account of beliefs, desires and intentions is not meant to bear any weight; I assume that Dretske has some such scheme in mind, and I mean to address what must be true of any such scheme.)

Now let us look at a particular case: at time t the fly buzzes and at $t + 8$ Dretske shoes. What has happened during 8? The fly causes a mediating perceptual event p , which causes the mediating belief-state b , which, along with desire-state d , causes the intention i , which causes Dretske's arm to move - something like that. In short, a lot of internal events happen, and all together they amount to his coming to think a fly is buzzing and then deciding to take some action, viz., shoeing. And in fact, given his learning history, all these mediating events and states mean what they ought to mean and cause what they ought to cause, but had any particular component been "wild" the sequence would have occurred just the same that's the way it is with actions: no locally potent meanings. Now Dretske says, in the postscript quoted above, that "one has to look at the meaning that was instrumental in shaping the behavior that is being explained," but no meaning was "instrumental" in shaping this particular action: it would have happened just as it happened, no matter what each element occurring during 8 meant or failed to mean. We do have room to advert to meanings if what we want to explain is how it comes to be that events of these (structural) types cause a behavior of this (structural) type at time $t + 8$. But that is also what we do when we account for the

etiological ancestry of reflexes. It is not that there is no differences at all between a reflex and an action. Of course there is. But it does not lie in the domain that Dretske explores.

Notes

- 1 This paper is an elaboration of my reply to Dretske at the SOFIA conference on information-based semantics at Tepoztlan, Mexico, August, 1988. The proceedings of that conference are published by Basil Blackwell in *Information, Semantics and Epistemology* (1990) edited by Enrique Villanueva. Dretske composed a postscript to his paper, published in that volume, responding to the objections I had raised in Tepoztlan, and the present paper responds in turn to that postscript.
- 2 It is interesting that here Dretske announces his allegiance to the same goal that motivates Fodor - the vindication of folk psychology taken neat - while dismissing Fodor's own solution as insufficiently responding to the folk-psychological intuition that (the Formality Condition be damned) meanings make it happen.
- 3 As Dretske's postscript makes clear, he did not intend to endorse natural selection in the paper under discussion. See the discussion of this postscript below.
- 4 See, for instance, my discussion (in "Mechanism and Responsibility," in *Essays on Freedom of Action* (London: Routledge and Kegan Paul, 1973), ed. T.Honderich, of the parallel case of the bogus belief (that he has an older brother living in Cleveland) surgically inserted into Tom (in *Brainstorms*, pp. 251-3); the discussion of the Panamanian debut of the two-bitser (how long does it take for the new functions to be "real"?): and my reply to Goldman, forthcoming in *Behavioral and Brain Sciences*.

References

- Dennett, D. 1969: *Content and Consciousness*. London: Routledge and Kegan Paul.
- 1971: *Intentional Systems*. *Journal of philosophy*. 8 (1971). pp. 8-106.
 - 1975: *Why the Law of Effect Will Not Go Away*. *Journal of the Theory of Social Behavior*, 5, p.p. 169-87.
 - 1978: *Brainstorms*. Cambridge. Mass: MIT Press/Bradford Books.
 - 1981: *The Kinds of Intentional Psychology*. In *Reduction, Time and Reality*. ed. R. Healey, Cambridge, Cambridge University Press.
 - 1983: *Intentional System in Cognitive Ethology: the Panglossian Paradigm Defended*. *Behavioral and Brain Sciences*. 6, pp. 34-90).
 - 1986. *The Logical Geography of Computational Approaches: A View from the East Pole*. In *Representation of Knowledge and Belief*, ed. M.Harnish M. Brand. University of Arizona Press.
 - 1987: *The Intentional Stance*. Cambridge: MIT Press/ Bradford Book.
 - 1988a: *When Philosophers Encounter Artificial Intelligence*. *Daedalus*, winter (issue on Artificial Intelligence), pp. 283-95.
 - 1988b: *Precis of The Intentional Stance*. *Behavioral and Brain Sciences*, II pp. 493-544.

- Dretske, Fred. 1981. Knowledge and the Flow of Information. Cambridge, Mass.: MIT Press/Bradford Book
- 1985: Machines and the Mental. Proceedings and Addresses of the American Philosophical Association, 59 (1985), p.p. 23-33.
- 1986: Misrepresentation. In Belief, ed. R. Bogdan. Oxford, Oxford University Press.
- Edekanm G. 1988. Neural Darwinism. New York: Basic Books.
- Haugeland, J. 1985. Artificial Intelligence: The Very Idea. (Cambridge, Mass.: MIT Press/Bradford Books.
- Schiffer. S. 1987: Remnants of Meaning. Cambridge, Mass. MIT Press/Bradford Books.
- Waltz, D. 1988. The Prospects for Building Truly Intelligent Machines. Deadalus winter (issue on artificial intelligence), pp. 191-212.