BEHAVIORAL AND BRAIN SCIENCES (1993) 16, 381–416
Printed in the United States of America

Continuing Commentary

Commentary on **Daniel C. Dennett (1988) Précis of *The Intentional Stance*. BBS 11:495–546.**

Abstract of the original article: The intentional stance is the strategy of prediction and explanation that attributes beliefs, desires, and other "intentional" states to systems – living and nonliving – and predicts future behavior from what it would be rational for an agent to do, given those beliefs and desires. Any system whose performance can be thus predicted and explained is an *intentional system*, whatever its innards. The strategy of treating parts of the world as intentional systems is the foundation of "folk psychology," but is also exploited (and is virtually unavoidable) in artificial intelligence and cognitive science more generally, as well as in evolutionary theory. An analysis of the role of the intentional stance and its presuppositions supports a naturalistic theory of mental states and events, their *content* or *intentionality*, and the relation between "mentalistic" levels of explanation and neurophysiological or mechanistic levels of explanation. As such, the analysis of the intentional stance grounds a theory of the mind and its relation to the body.

Comments on Dennett from a cautious ally

Jonathan Bennett
Philosophy Department, Syracuse University, Syracuse, NY 13244
Electronic mail: *bennett@suvm.syr.edu*

1. Indeterminacy of content. In this commentary, unadorned page numbers under 350 refer to Dennett (1987) – *The intentional stance* (hereafter referred to as *Stance*); and those over 495 refer to Dennett (1988) – mostly to material by him but occasionally to remarks of his critics. Although this commentary will focus on disagreements, I should say now that I am in Dennett's camp and am deeply in debt to his work in the philosophy of mind, which I think is wider, deeper, more various, and more fruitful than mine or anyone else's. Still, I have some ideas and emphases that I think he could profit from.

In the final chapter of *Stance*, Dennett compares his work with that of several others, including me. He sees me as having a position like his, the main difference being that I think (as he does not) that our attributions of mental content can always be highly determinate (pp. 347ff). There are in fact differences between us, but this isn't one of them. I want to get this straight, so as to clear the decks for the positive points I am going to make.

There is some indeterminacy and there could be lots of it; Dennett's case for that is unanswerable. As for how much there actually is: I don't know and don't even suspect; there is simply no declared issue between Dennett and myself on that. Nor do we disagree on a related matter. If there is no evidence that settles whether the animal believes that *P* or believes that *Q*, should we say that nevertheless one of these is right and it's just that we can't know which it is? Dennett says no. I perfectly agree.

I have argued against Dennett on the matter of how we get from premises about behavior to conclusions about thoughts (Bennett 1983; see also sect. 3 below). He seems to represent the process as free-ranging, somewhat haphazard, a matter of guesses and luck that is subject to only two extremely mild constraints whereas I contend that there is or can be a good deal of discipline to it, that there are fairly definite conceptual structures that can guide us in deciding what mentalistic attributions are supported by what facts about behavior. But that disagreement between us has nothing to do with how determinate the attributions of content can be. The latter question is a matter of relative detail, to be settled by understanding the conceptual structure and studying the animal behavior in the light of it.

Presumably someone who believes that thoughts and wants must be highly determinate will be led to reject Quine's (1960) thesis about the indeterminacy of translation: The determinate Gricean wants of the speaker, he will think, must generate a determinate answer to the question "What did he mean by what he uttered?" But not conversely. One may disagree with Quine's thesis that the meanings in any language must involve much indeterminacy, as I do, without holding that thoughts are all determinate.

Dennett thinks that I believe in determinate content because I reject Quine's thesis about the indeterminacy of translation. This seems to me to be a philosophical mistake: There is no inconsistency in rejecting Quine's thesis about *language* while believing that *thoughts* are in general not very determinate. One of the things that makes this possible is the fact that sentences, unlike thoughts, have separately meaningful parts. In a way described in Bennett (1976) and Blackburn (1975), that creates a possibility for determinateness in linguistic meaning without relying on determinateness of thought. Quine's thesis has almost nothing to do with any serious interest of Dennett's. The only link is the fact that whoever disagrees with Dennett about the indeterminacy of thoughts will also disagree with Quine. In setting himself against everyone who disagrees with Quine, Dennett is multiplying opponents beyond necessity. I think he does that rather a lot.

There are three important respects in which I do part company from Dennett's views or procedures. I shall give two of them a section each, and then turn to the third.

2. The unity condition. We can adopt the intentional stance toward a thermostat, Dennett says. At noon the thermostat *closed the switch* because it *perceived* that the temperature was below 65°, *wanted* it to be at 65°, and *thought* that closing the switch was the way to raise it. There we see the four elements of a functionalist theory of mind: Perceptual inputs and behavioral outputs are interpreted in terms of a psychology of beliefs and desires. Dennett's penchant for saying things like this, illustrating his idea that intentionality is rooted in a *stance* that we are free to adopt or not as we choose, has been criticized, often intemperately. I shall offer a cooler criticism, based on Bennett (1976, sect. 21 and 22).

Since chemical explanations involve principles that go wider and deeper and theoretically admit of greater precision than intentionalist ones, why should they not always be preferred? There are four answers one might give. (1) Some human movements cannot be explained chemically but can be explained in

 0140-525X/93 $5.00+.00

terms of thoughts and wants. (2) We often don't know the chemical explanation, which entitles us to use an intentional one, *faute de mieux*. (3) Justification is not needed; there are no constraints on our choice of how to look at the animal and what concepts to apply to it. (The fourth answer will be highlighted shortly.)

Nobody today believes (1). Dennett sometimes has recourse to (2), as on p. 315; but much of what he writes sounds like (3), which brings thermostats smoothly into the story and makes some people's blood boil. I think he needs answer (4), which is as follows:

(4) An intentionalist explanation of behavior brings out patterns, provides groupings and comparisons, that a chemical one would miss. What the animal did belongs to a class of behaviors in which it wants food and does what it thinks will provide food, and there is no unitary chemical explanation that covers just this range of data. This animal seeks food in many different ways, triggered by different sensory inputs, and it is not credible that a mechanistic, physiological view of the facts will reveal any unity in them that they do not share with behaviors that were not food-seeking at all. If this unifying view of the facts answers to our interests, gives us one kind of understanding of the animal, and facilitates predictions of a kind that are otherwise impossible (predictions like "It will go after that rabbit somehow"), we have reasons for adopting it, while still acknowledging that each of the explained episodes, taken separately, could be explained in a way that is deeper and broader and – other things being equal – preferable.

The main thrust of this is something that Dennett himself has presented more clearly and eloquently than anyone (for example, on pp. 22ff), but he doesn't properly use it to justify the intentional stance. That is, he doesn't make it a matter of doctrine that intentional concepts are legitimate only when they satisfy the "unity condition," as I call it, that is, only when they conceptually unify episodes that are otherwise disparate. That doctrine would save him from skidding down to where the thermostats are. The abstract analogy between what thermostats do and how thinking animals behave is real, and worth pointing out. But people's sense that it is just wrong to talk in the thought-want way about thermostats could be explained by their being sure that the facts about thermostats rule out justification (4).

(This makes the justification of intentionality a matter of degree, but with any matter of degree there can be things that fall right off the bottom end of the scale. Also, what I am saying has nothing to do with the difference between animals and artifacts; it is only about degree and kind of complexity of behavior patterns.)

When Dennett writes: "Nothing without a great deal of structural and processing complexity could conceivably realize an intentional system of any interest" (p. 60), I would replace that last phrase by "a genuinely intentional system," leaving "interest" out of it. Much of the time, indeed, that seems to be Dennett's own view. In a case of "zero-order intentionality," he writes, "what had seemed at first to be explicable in terms of belief and desire turns out to have a deflated interpretation" (p. 539). Again: "If one gets confirmation of a much too simple mechanical explanation . . . , this really does disconfirm the fancy intentional level account" (pp. 542–43). It looks as though Dennett is here relying on the unity condition as a mark of the intentional, but he does not ever make this the matter of explicit doctrine it deserves to be.

As soon as we have some conditions that a thing must meet if we are to be justified in interpreting it intentionally, we can demote the notion of the intentional "stance." This gives so many people so much trouble that I can't help thinking that Dennett would do well to drop it. He does say that when we attribute thoughts and wants to something, there is a fact of the matter regarding whether the thing's behavior manifests patterns of the right kind (p. 24); but he refuses to drop the "stance" language, though it inevitably suggests that he is less of a realist and more of a libertine about intentionality than he really is.

While I am on the "realism" theme, I have a suggestion that I hope might help. Rereading Dennett (1988), I am struck by how often his critics put the question in terms of realism about "beliefs and desires"; see, for example, Dretske (1988a) and Stich (1988). It is worth separating two questions. (1) How realistic should we be about statements of the form "*x* believes that *P*" and "*x* wants it to be the case that *Q*"? (2) Are there really any such items as beliefs and desires? Even if attributions in the language of "believes that" and "wants" are perfectly solid, the nouns "belief" and "desire" might be misleading *façons de parler*.

3. Getting from behavior to mental content. The unity requirement is relevant not only to *whether* but also to *how* one is entitled to bring intentional concepts to bear on a creature. In his seventh chapter Dennett offers two rules guiding the interpretation of the doings of animals in intentional terms. (1) In the absence of evidence to the contrary, assume that the animal does what it thinks will produce what it wants. That is so right that it is analytic. The innermost core of the functionalist approach to intentionality – the Ur springboard for the concepts of belief and desire – is the explanation of an animal's doing *A* through the hypotheses that it wants *G* and that it thinks that doing *A* is a way to get it. (2) Don't inflate; that is, don't attribute mentality when the facts can as well be explained without it, and don't attribute complex or high-level thoughts when the facts can be explained just as well by postulating simpler or lower-level ones. That is good too; it is in fact a use of the unity condition or something like it.

It would, however, be a poor outlook if those two were our only guides.[1] If we want help in devising mentalistic hypotheses to explain animal behavior, (1) does not provide it; it says that attributed beliefs and desires must relate to one another in a certain way, but offers no other constraints. Dennett says somewhere that we can get started by hypothesizing that the animal thinks and wants what *we* would be thinking and wanting if we behaved like that. That would indeed be a start, but how do we get from it to something better? Not always through (2) alone, because our first try might be wrong for reasons other than that it is inflated. Furthermore, it would be good to have help in applying (2), which is not as straightforward as it might seem.

When an interpretation is "wrung from the exploitation of the intentional stance" (Dennett, p. 312), the procedure need not be one of flailing around until we get lucky. Here is a partial sketch of something that would provide more leverage on problems of interpretation than Dennett's two rules can do unaided. I mainly want to illustrate the *kind* of thing I mean by "structure" in the move from behavior to mentalistic interpretation. Even if this particular proposal fails, the general point may still stand.

To bring my one point into sharp focus, I shall idealize: I consider an animal whose goals or standing desires do not change, and which is free from practical conflict. Its beliefs never bring into play two desires that cannot both be satisfied.[2] Now let us consider the situations in which this animal thinks there is something it can do that will satisfy one of its desires. Let us take, specifically, the class of behavioral episodes in which we suspect that it aims to get food.

Cognitive explanations are not supported if the relevant behavior is all covered by this:

> Whenever the animal picks up a trace of chemical *C* in the water, it waves its tentacles and then brings them toward its mouth.

That plainly invites explanation in terms of simple stimulus-response triggers, giving no purchase to explanation in terms of wants and thoughts. This is clearly implied by the unity condition, which will not let us adopt an intentional explanation if the facts are adequately caught in the statement that whenever the animal has this *stimulus*-kind of input, it produces that *motor*-kind of output.

For an intentional account to be honest, we need something

more like this: A behavioral pattern involving (1) a class of environments whose best-unified description is that in each of them there is something the animal can do that will bring it food, and (2) a class of outputs that are united only in that with each of them the animal moves in some way that results in its getting food. That is only a first approximation, however. It would be right only if our animal never went wrong about what would bring it food, and that is idealizing too far: The possibility of error is too important to be set aside, even in the initial stages of the conceptual story (see sect. 5 below).

So we need to replace that account of the class of inputs by something like this:

> Given the animal's perceptual apparatus (its "quality space," etc.), each of the relevant environments is significantly similar to those in which there is something the animal can do that will bring food.

I shall designate as "the comparison set" for a given behavioral episode *A* the class of environments that (1) are relevantly similar to the one in which *A* occurs and (2) are such that in each of them there really is something the animal can do that will bring it food. Then I can give an amended description of the outputs, namely:

> On each occasion, the animal moves in a way that would be likely to bring it food if the environment were a member of the comparison set.

Of course in most cases the actual environment *is* a member of the comparison set.

Now, both versions of the input side of the story involve the notion of food-getting behavior: In the simple version, each environment is one where *the animal can get food;* in the version that allows for error, each environment is significantly like ones in which *the animal can get food.* The notion of the animal's getting food cannot be replaced by anything unitary that does not involve that notion; this is why it is legitimate to explain these behavioral episodes in terms of the animal's thinking that what it is doing will get it food. If there were some single stimulus-kind of sensory input – a particular kind of patch in its visual field, a particular kind of smell, or the like – so that on each relevant occasion the animal receives a stimulus of that kind, then these behaviors fail to support the attribution of wants and thoughts about getting food. The getting-food content is justified by the need for the notion of food-getting in characterizing the class of environments in which the behavior occurs.

An analogous story can be told on the output side. A class of behaviors that do not belong to any one motor kind may be united by each being of some kind that usually leads to the ingestion of food. For more details (of which there are plenty), see Bennett (1991b, pp. 46–48).

My guiding rule applies not only to whether it is all right to attribute content, but also to what content to attribute. Did the monkey want its companions to *believe there was a leopard nearby* or merely to *climb a tree?* Evidence that the former attribution is right requires a class of behaviors in which it is not always the case that the animal's behavior is apt to get its companions to climb trees, but it is always the case that its behavior is apt to get them to think there is a leopard nearby. [See BBS multiple book review of Cheney & Seyfarth's "How Monkeys See the World" *BBS* 15(1) 1992.]

How, in the absence of language, could that be? Well, if the monkeys can use the information that a leopard is nearby in various ways, and the subject animal's warning cries occur when any one of these uses could be made of the information, the relevant class of environments is unified as follows:

> The environment is such that the subject animal can behave in a manner that will get its companions to *behave in a manner appropriate to the information that there is a leopard nearby.*

Just as in the earlier example, the class of environments is unified with help from the concept of food-getting, so here the class of environments is unified with help from the concept of behaving in a manner appropriate to the information that there is a leopard nearby; so we are entitled to put *that* into the animal's desire and belief. It will have to pass muster for the animal's wanting to get the others to *believe* there is a leopard nearby: This is as near as we can get in the absence of language (see Bennett 1991a for details).

This sketches part of the story, showing the kind of thing I mean when I say that Dennett doesn't do justice to the disciplined conceptual structure in the relationship between behavior and mental content. This is not, incidentally, to imply that such attributions are usually or always highly determinate. Rather, it suggests Dennett-like reasons why they often cannot be.

4. Intentionality and evolution. In the eighth chapter of *Stance,* Dennett connects intentionality with evolution, for which he is taken to task by several of his critics. I agree with him that the intentionality of individuals is abstractly *similar* to what goes on in evolution, which is why we can coherently talk about "the designs and plans of Mother Nature."[3] (I am inclined therefore to disagree with what seems to be Ringen's main thesis in his accompanying commentary.)

Dennett also alleges that there is a conceptual link between intentionality and evolution. I think there is too, as I shall explain in section 7 below, but not the link that Dennett argues for. He contends that we cannot get any notion of what an animal thinks except with help from facts about what it is designed to do, the "design" in question being that of evolution: "It is only relative to . . . design 'choices' or evolution-'endorsed' purposes . . . that we can identify behaviors, actions, perceptions, beliefs, or any of the other categories of folk psychology" (p. 300). Also: "Attributions of intentional states to us cannot be sustained . . . without appeal to assumptions about 'what Mother Nature had in mind'" (p. 314).

(Dennett insists that even when we bring Mother Nature into the picture, we won't get fully determinate content, and he uses this as a principle of inference: "is not independent of the intentions and purposes of Mother Nature, and *hence* is . . . subject to indeterminacy of interpretation" [p. 305, emphasis added]; "you are . . . just a product of natural selection, whose intentionality is thus derivative and *hence* potentially indeterminate" [p. 313, emphasis added]. I don't know what the warrant is for these uses of "hence.")

If this were right, what could explain the existence of folk psychology in the centuries before evolutionary theory was thought of? In fact, we can have reason to attribute thoughts and wants to an animal without implying anything about how it got to be that way or about why there are such animals. What counts is how the animal does (would) relate to actual (possible) environments: That is the behavioral ground in which our concept of cognition takes root. Dennett thinks this has to be supplemented, but why?

In Dennett's defense of this position, two things are going on. One is an attack on "intrinsic intentionality," that is, on the view that the facts about what if anything an animal believes and wants are monadic facts about it; they could in principle be established by attending to that animal and nothing else at all.[4]

Now, there are two standpoints from which one might deny this. (1) Intentionality *conceptually involves an animal's relations to its environment.* Many facts about how the animal did, will, and would move, in what kinds of environment, and with which upshot, are relevant to what beliefs and desires explain something that it is doing right now. Relations to context are relevant to what the animal believes and wants, not merely to our evidence about what it believes and wants. Dennett accepts this, and so do I; I will have no quarrel with this here. (2) *Any instance of intentionality must be derived from the (near-) intentionality of something else.* Thus, according to Dennett, a certain machine can be described as having certain "desires" and "beliefs" only on the strength of facts about what its owners want it for; and, again according to him, animals count as having desires and beliefs only on the strength of facts about the "plans" of Mother Nature (or, less poetically, facts about what the

animals' various features were selected for). This form of the rejection of intrinsic intentionality leads to the thesis now under discussion, that animal intentionality is conceptually parasitic on facts about evolution.

Dennett argues against philosophers of mind who reject both (1) and (2) because they believe in intrinsic intentionality. After many readings of his eighth chapter, I still think that some of his confidence in (2) comes from his being so sure that *these* opponents of it are wrong. Now, although the point didn't come through clearly in the commentaries on Dennett (1988), there must be many of us philosophers who agree with Dennett about (1) and are not convinced about (2). What we need is a defense that keeps quiet about "intrinsic intentionality" and homes in on what is special to (2).

5. Grounds for attributing error. Dennett does present such a defense, by pointing to a kind of problem that can arise when one is trying to settle what an animal believes and wants on the basis of its relations to its environments – a problem he says can be solved by appeal to facts about evolution and, he evidently thinks, in no other way.

When presented with certain kinds of sensory stimuli, an animal makes certain kinds of movements; the usual result is that it captures and eats a fly, but sometimes it takes in a piece of bark about the size and shape of a fly that is blown in front of it by the wind. Let us pretend that we have here enough of the right sort of complexity to justify the interpretation in terms of beliefs and desires (see sect. 2 above); the question is, What beliefs and desires shall we attribute to this animal? We could say that (a) the animal always thinks it is getting a fly and that sometimes it is wrong about this, or that (b) when it gets a fly it thinks it is getting a fly, and when it gets bark it thinks it is getting bark, or that (c) it always thinks it is getting either a fly or a piece of bark, or that (d) it always thinks that it is getting something that is small and dark; and there are other possibilities too. The crucial question Dennett raises is: What resources do we have for preferring (a)? If there cannot be a basis for this in some such cases, we are in trouble. This is not because of any facts about this animal and this behavior in particular, nor is it because (c) and (d) are unacceptable because they are somewhat "indeterminate." The crucial point is that a viable system of intentionality *must have a basis for sometimes crediting the subject with false beliefs*, and in our present example (a) is the only diagnosis that attributes error. Dennett is right: If we can never attribute error, we don't have a viable system of intentional concepts.[5]

Dennett says that we may be able to choose (a) on the strength of the plans of Mother Nature. The question is: Why did the evolutionary process select the pattern of behavior we are trying to interpret? If it was selected because it generally leads to the ingesting of flies, that gives us a nudge toward (a), that is, a basis for saying that in all these behaviors the animal thinks (sometimes wrongly) that it is going to get a fly.

That is indeed one way of filling a gap in our intentionalist story. There are others – for example, one based on the facts about which outcomes of the behavior are conducive to the animal's survival and health – and Dennett says nothing about why we must have his rather than one of the others. But I shall not pursue this line of argument, because the issue it concerns is peripheral and minor. The important point is that we can have grounds for attributing error without appealing to facts about evolution or any alternative, such as facts about health and survival. We can "wring from the intentional stance" itself – looking only at the animal's relations to its environments – reasonably grounded interpretations, some of which say that "the animal wrongly believed that *P*."

In Bennett (1976, Ch. 2–4), I offer some proposals for how to do this. Even if they fail entirely (which I doubt), they at least create a presumption that the job can be done, and that there is no absolute need for either evolution or health as bases for attributing error. I shall merely sketch the core idea.

We should look at our subject animal's other engagement with flies and with small pieces of bark. If it has other patterns of behavior that also lead to its eating flies, but passes up all other chances to eat bark, that is evidence that it has the eating of flies but not of bark as a fairly permanent goal, and thus that, in all the behavior we were initially concerned with, it thought (sometimes wrongly) that it was getting a fly. The whole story is more complicated than this, but this is enough to be going on with.

"What if the animal has no other encounters with either flies or bark?" Then perhaps this behavior pattern ought not to be handled in intentional terms at all (see sect. 2 above). However, if there are enough complexities in other parts of its behavior to justify crediting it with beliefs and desires, we may well think it is reasonable to apply the intentionality apparatus to its fly-bark behavior as well. *Then* what are we to do about becoming entitled to attribute error? Well, we could appeal to facts about survival and health, or to facts about evolution; or we could say we are not entitled to attribute error in any of these cases, and that on each occasion the animal correctly believes that it is capturing something small and dark; or we could say that the behavior probably warrants being handled in intentional terms, but we cannot decide how. It doesn't matter much: We are only discussing whether to attribute error in this one little corner of the animal behavior kingdom; we no longer have at stake the possibility of attributing error at all, anywhere. The pressure is off.

Dennett: "Attributions of intentional states to us cannot be sustained . . . without appeal to assumptions about 'what Mother Nature had in mind'" (p. 314). I claim to have shown that this is not so.

6. The vending machine. Dennett leads into his discussion of error through a story about a vending machine that will accept two sorts of coin, U.S. quarters and Panamanian quarter-balboas; after years of use in the United States the machine gets years of use in Panama, and Dennett asks: When and why does the machine switch from being one for which quarters are coins and quarter-balboas are slugs to being one for which the reverse holds? This is put in terms of the intentional stance. Initially we are to see the machine as (so to speak) thinking that it is getting quarters, and sometimes wrongly thinking this about a Panamanian quarter-balboa. After years of use in a Panamanian bar, it seem rather to be (as it were) thinking that it is getting quarter-balboas, and sometimes wrongly thinking this about a U.S. quarter. What makes the difference? Not the relative frequency of the two sorts of coins, says Dennett, but rather the wishes of the machine's owners. This is offered as isomorphic with the way in which, according to Dennett, the content of our thoughts is derived from facts about the plans of Mother Nature.

Because this vending machine example dominates the chapter, it is worthwhile to point out three flaws in it, of which the second is probably the most serious.

(1). The analogy is a bad one right on the surface. Facts about the plans of Mother Nature are facts about the origins of animals and their behavior patterns: Facts about what the vending machine's owners want it to do are not about its origins. That is why what counts as veridical for it at one time counts as erroneous later; Dennett makes no provision for that in the case of any individual animal, and could not do so on the basis of facts about evolution.

(2). The vending machine's behavior is too simple to illustrate Dennett's point. If we do not appeal to the owners' wishes, then indeed we have no solid basis for saying which of its states are "errors"; but what of it? The fact that we cannot make a concept of error work in this case, where the item is so far from satisfying the unity condition, simply doesn't matter. We do not have to solve the "problem of error" for the vending machine: We can say "There is no 'error' here" without being at risk of having to say that there is no error in behaviorally complex animals.

Point (2) has nothing to do with the fact that the machine is an artifact, or that it is a purely physical system; and it owes nothing

to the assumption that intentionality must be intrinsic. When people object to the vending machine example, Dennett rather quickly suspects the worst of them; I'm trying to bring out that there are definite, decisive, *limited* grounds for objection.

(3). In saying that the vending machine offers such a low-level analogue of intentionality that it doesn't do the work that Dennett demands of it, I was conceding too much. The vending machine presents no analogue of intentionality, however primitive. The skeleton of intentionality is this: A certain object does *A* because it wants *G* to obtain and thinks that doing *A* is the way to make *G* obtain. That form of teleological generalization is the nonnegotiable minimum that is needed for even a simulacrum of intentionality. That bare skeleton is exhibited by thermostats but not by ordinary vending machines. What is the vending machine's goal *G*? Ingesting coins? That was Ned Block's answer when he first brought vending machines into this literature (Block 1980, p. 173), but it is absurd. The machine does not do anything that gets coins into it; there is no value of *A* such that whenever the machine "thinks" that doing *A* is a way to get a coin into it, it does *A*. Well, then, getting rid of bottles of drink? If so, then each kind of coin is always a means to the machine's goal, and there is no purchase for the notion of error. To keep his example in business, Dennett must say that the machine's goal is pleasing its owners. But that would be frivolous, and not like his usual way – which shocks some, but is perfectly sober – of illustrating the structure of intentionality through low-level examples.

Still, one could replace the vending machine example by one that has the right formal properties, and then my first and second objections would still stand and would shine out more clearly. In the preceding paragraph I was merely trying to get mud out of the way.

7. A real place for evolution. Despite my criticisms of Dennett's way of forcing evolution into the intentionality picture, I do think it has a place there, for a reason that I now explain.

It starts from the fact that attributions of beliefs and desires are nothing unless they help to explain behavior. The intentional concepts that we ordinarily use, the ones that are defined by folk psychology, simply *are* explanatory; and attempts to analyze them come to grief if they don't give a central place to that fact.[6] Now, what explains must predict, and for that we need generalizations – most basically the teleological ones that I stressed in the third part of section 6 above. But to support predictions, a generalization must be projectible, that is, true because its antecedent is reliably linked with its consequent. In Bennett (1976) I completely overlooked this. I stressed the explanatoriness of the concepts in question, but in my own account of them I backed them with generalizations which might, for all I said to the contrary, have been accidental. The gap is filled in Bennett (1991c); I shall sketch the point here.

There is no problem about explaining or predicting an animal's going from a stimulus of sensory kind *S* to a movement of motor kind *M* if it has done this often enough to convince us that it has some structurally grounded disposition to link this kind of input with that kind of output. That link, however, corresponds to a single mechanism; an explanation that exploits it does not satisfy the unity condition, and so does not involve intentionality and thus falls outside the scope of our present question. We want to know: What can make it all right for us to trust an intentional or teleological generalization to lead us from some *S-M* linkages to predict *others*? We have observed our animal in a range of situations where there is evidence that it can get food, and it has usually done a food-getting thing. The evidence has consisted (variously) in this or that smell, this or that sound, and so on; the resultant movements have involved running, swimming, climbing, and killing. Now it is in the presence of different sensory evidence that food is available, and to get the food it would have to dig. What would entitle us, even weakly, to regard its previously observed food-getting activities as evidence that it will now dig? Setting aside special creation, one form of which would do the job, two answers remain.

(1) First there is evolution. Of all the potential mechanisms that were awarded a try-out in the animal's ancestors, relatively few were taken in permanently; among the survivors were the bunch of mechanisms that make their owner a food-getter, and *that is why they survived.* Why does this animal contain a lot of mechanisms that make it a food-getter? It inherited them from a gene pool that contained them *because they make their owner a food-getter.* That makes it more than a coincidence that the animal has many mechanisms that are united in their food-getting tendency, and lays a clear basis for predictions and explanations that bring in intentionality.

This is quite different from what Dennett has been saying about what mental content to attribute to the animal. Rather, it starts at the point where we have decided what content (if any) to attribute, and addresses the question of whether this attribution supports predictions and explanations. Also, it is not the only solution to the problem that it addresses, as I now show.

(2) Second, there is individual learning. If an animal is disposed to food-getting and is educable about this, that fact alone entitles us to predict and explain some of its behavior. Its having food as a goal, together with its being able to learn from experience which movements yield which results in which circumstances, jointly give us reason to predict that it will pursue food in ways and on clues that we have not previously seen it use. The animal itself must have encountered the clues and tried the movements, but even if we have not seen it do so, we can reasonably guess that it has experienced the relevant input-output link and will therefore dig for food on the present occasion.

In answer (1) the individual animal was not credited with being able to modify its set of *S-M* linkages in the light of good or bad experience. For all my account implied to the contrary, what evolved might have been a perfectly rigid set of *S-M* linkages which had the effect, in a certain kind of world, of making its possessors food-getters. I'll bet that there are no such animals,[7] but in theory there could be. Nor is there any conceptual confusion in the idea, admittedly a biologically lunatic one, of an animal that could learn but did not evolve. So the two ideas are independent, which means that we have two sources for the power of intentional hypotheses about animals to predict and explain.

In the eighth chapter of *Stance,* Dennett argues at length against Dretske's (1985; 1986) attempt to found intentionality on learning, which he sees as a rival to his attempt to found it on evolution. What I have argued in this section throws an odd light on that debate. I contend that intentionality is founded on the disjunction of learning and evolution.

NOTES

1. The following remarks improve on my poorly expressed and inadequately thought out comments on this material of Dennett's in Bennett (1983).

2. For suggestions about how to remove these simplifications and get nearer to real animal behavior, see Bennett (1976, sects. 18–20).

3. Dennett (1987, p. 299). I comment on this similarity in Bennett (1976, pp. 78f, 204–10).

4. Dennett seems to assume that a believer in intrinsic intentionality will also think that *what* the animal believes and wants is always highly determinate (see, for example, the paragraph on pp. 303f). Why? Couldn't even a strict Cartesian think that some thoughts are vague? The notion of (in)determinate mental content keeps cropping up in the chapter; it is allowed to hook into everything, often in ways I do not understand.

5. I argue for this in fragmentary ways in Bennett (1976) and elsewhere. Here is a start: If *x* is subject neither to error nor to ignorance, then "*x* believes that . . . " is vacuous, does no work, and is equivalent to "It is true that . . . ."

6. For some examples, see Bennett (1976, sect. 13).

7. For reasons given in Dennett (1974).

Dennett's intentions and Darwin's legacy

Jon Ringen

Department of Philosophy, Indiana University South Bend, South Bend, IN 46634

Electronic mail: *jringen@indiana.edu*

The intentional stance (Dennett 1987, henceforth *Stance;* Dennett 1988, henceforth *Précis*) highlights the conceptual tensions introduced by Darwin's (1923) comparison between natural and artificial selection. In *Stance*, the tensions arise in Dennett's account of two important insights: (1) that adaptationism in biology and cognitivism in psychology embody similar intentionalistic and teleological modes of explanation (*Stance*, p. 282ff; *Précis*, p. 502), and (2) that natural selection provides a promising model of explanation for both the phenomena of adaptedness, which are adaptationists' central concern, and the phenomena of concept use and acquisition and rational choice, which concern cognitivists (*Stance*, p. 278ff, p. 316ff; *Précis*, pp. 502, 503). Dennett embraces both adaptationism and cognitivism and endorses (*Stance*, p. 348) the views on intentionalistic and teleological explanation presented by Bennett (1976). Dennett also argues that, as a model for scientific explanation, natural selection is intimately linked with the type of mentalistic explanation used by modern cognitivists and the type of functional explanation used by modern adaptationists. In sum, Dennett argues that his two insights are complementary. My comments are intended to show that as articulated in *Stance* and the *BBS Précis*, Dennett's insights are in serious tension: Dennett must choose between two conceptions of teleology and intention, one that can be modeled on natural selection and one (namely Bennett's) that cannot.

The issue here is not merely whether Bennett's views and Dennett's differ. The underlying issue is, in fact, one that must be faced by both the defenders and critics of adaptationism in biology and the defenders and critics of cognitivism in psychology: Is final causation (in Aristotle's sense) indispensable in modern biology and psychology? The issue is posed by Dennett's claim (e.g., *Stance*, pp. 316, 318) that the biologist's notion of "selection *for* a characteristic" is an inherently intentional notion in exactly the sense of "intentional" which the intentional stance embodies. Dennett's endorsement of Bennett (and of Aristotelian "why questions," *Stance*, pp. 238, 286) presents the problems about teleology that biologists, psychologists, and philosophers must address. There are two related sets of issues to consider, one conceptual and the other practical.

Dennett suggests (*Précis*, pp. 502, 541; *Stance*, pp. 314–20) that explanations in terms of natural selection embody a distinction between "selection *for* a characteristic" and "selection *of* a characteristic" which cannot be maintained if mentalistic and teleological notions are given up. There are prima facie reasons for contesting this. In particular, a clear conceptual distinction between "selection *of*" and "selection *for*" can be made solely in terms of differences in efficient causal pathways. Elliot Sober (1984; 1986) provides such a nonteleological characterization of the distinction. Roughly speaking, a characteristic is selected *for* just when it is a causal factor in its bearer's actual reproductive success. There may be (as in pleiotropy) selection *of* characteristics which are not selected *for* when characteristics of reproducing organisms are transmitted to the next generation (as they frequently are) even though they decrease or are neutral with respect to their bearer's chances of reproductive success (for a useful discussion see Shanahan 1990). There are practical difficulties involved in determining whether a particular (genotypic, phenotypic, or group) characteristic was selected *for* or even selected, but there is no need, in principle, to bring in notions that are mentalistic or teleological in any sense (including Bennett's) to explicate or apply this distinction. Any suggestion that such notions are required just seems to be false.

Dennett's thesis about the intimate connection between "intention" and "selection *for*" would be defensible if his conception of teleology and intention could be explicated in terms of biologists' notion of "selection *for*." Such a conception of teleology and intention has many proponents, but Bennett cannot be among them. Seeing why this is so clarifies the significance of treating natural selection as an optimization process.

According to Bennett's account, teleological principles make the phenomena which conform to them a function of whatever is, under prevailing circumstances, causally appropriate (e.g., causally necessary and causally sufficient) for maintaining or bringing about a specific state of affairs. Such principles entail that when circumstances change what is causally appropriate for the (goal) state specified by the principle (in Bennett's [1976, p. 38] terminology, when the relevant *instrumental properties* change), then whatever is (within the system's "repertoire" and "registered as") causally appropriate under the new circumstances will occur. Thus, teleological principles provide a basis for predicting what response to new circumstances a system which conforms to them will produce. The main principles that define the process of natural selection do not provide a basis for such predictions. Assessing the significance of this fact is a complicated matter (see, e.g., Dupré 1987) that depends as much on how "selection process" is defined as on what processes of (evolutionary) change actually occur. Nevertheless, it is possible to consider the conceptual issues involved.

Natural selection involves two processes: generation of variation and selection from among the variants generated (see, e.g., Darden & Cain 1989; Staddon 1983). According to the usual account, neither of these processes is, by itself, capable of generating responses to changing circumstances that match the predictions of principles that define teleology in Bennett's sense. Consider the process Dennett emphasizes: Selection *for* is a sorting process (Amundson 1989). Sorting processes do exhibit features reminiscent of intentionality. They "register" or "discriminate" the features they "select *for*." However, sorting processes only operate on characteristics already produced, so by themselves they cannot provide a basis for predicting what characteristics will emerge when instrumental properties change. The point can be put as follows: When environmental changes alter what is required to survive and reproduce, the principles defining sorting processes such as selection *for* only provide a basis for conditional predictions of the form: *If* a fitness-enhancing trait emerges, it will be selected *for*. The principles do not provide a basis for any *categorical* prediction that fitness-enhancing traits will emerge. Such (categorical) predictions, however, are the kind that Bennett's teleological principles license. This establishes that processes that are teleological in Bennett's sense cannot be identified with the process of selection *for*. Selection *for* is a *sorting* process; it does not *generate* variation.

This does *not* settle the issue of whether Bennett's conception of teleology can be explicated in terms of the principles that specify how selection *for* operates in natural selection. According to the customary view, selection *for* operates in tandem with processes that generate variation. Minimally, Bennett's teleological processes require the latter. It is a conceptual possibility that the processes of generation involved in natural selection are goal-directed (i.e., conform to principles that are teleological) in exactly Bennett's sense. If this were so, then when (critical) changes in a system's environment changed what was appropriate for reproductive success, the characteristics generated would tend to be those appropriate for reproductive success in the new environment. This would support the contention that Bennett's concept of teleology can be explicated in terms of principles governing natural selection, but it would also render Darwin's legacy moot. Goal-directed processes of generation would make sorting processes like selection *for* largely irrelevant. Accordingly, in standard models of natural selection (see Gould 1980; Maull 1977) the processes of generation are random, stochastic, or deterministic in ways that are explicitly contrasted with teleology, agency, goal direction, or inten-

tionality. Standard conceptions of processes of generation, like the standard conceptions of "selection *for*," cannot provide a basis for defining processes that are goal-directed in Bennett's sense.

Attempts to explicate Bennett's conception of teleology and intention in relation to standard conceptions of natural selection cannot focus on either selection or generation alone. They must begin with the conception of nonteleological processes of generation *interacting* with nonteleological processes of selection *for*. Dennett would surely agree with this judgment, but then he must give an account of how this can be done. No such account has yet been provided by Dennett. Dennett does (correctly) point out (*Précis*, p. 542) that the *effects* of selection processes are sometimes the same as the *effects* teleological processes would produce: Both types of process may leave characteristics that are adapted to existing circumstances in the sense that they are (under prevailing circumstances) causally appropriate for enhancing reproduction. But this coincidence of effects is not enough to establish that the process of natural selection is teleological in Bennett's sense. It also needs to be shown that nonteleological principles of generation and sorting provide the basis for categorical predictions of the sort Bennett's teleological principles provide. Showing this would ground Dennett's endorsement of Bennett. It would also ground a decidedly Panglossian view of natural selection as a process that generates just those characteristics that changing circumstances make instrumental for the goal of reproductive success. It is reasonable to doubt that nonteleological processes of generation and sorting *necessarily* have such effects.

As far as I can see, these kinds of considerations settle the conceptual issues raised by the suggestion that "selection *for*" is intentional (and hence teleological) in the sense specified by Bennett. "Selection *for*" and Bennett's conception of teleology and intention are independent in the senses indicated in my arguments. Still, there is a broader *practical* question that motivates Dennett's discussion: Can the biologists' notion be *applied* independently of the framework provided by Bennett's notions of "intention," "goal direction," and "teleology"? This question raises issues concerning the nature of optimality models and their role in adaptationist thinking. [See also Schoemaker: "The Quest for Optimality" *BBS* 14(2) 1991 and Anderson "Is Human Cognition Adaptive?" *BBS* 14(3) 1991.]

Biologists encounter enormous practical problems in determining which traits of a system were selected *for*. Here, adaptationists frequently use "optimality models" as a "shortcut" or "heuristic" (Beatty 1980; Lewontin 1979) for identifying those traits (adaptations) that have been selected *for* in the course of biological evolution. An example is provided by Holling's (1964) investigation of optimal food size. Holling argues that given the physical structure of mantid forelegs, mantids cannot grasp prey larger than a certain size. He also argues that considerations of energy efficiency dictate that mantids not waste time or energy on prey smaller than that size. This is an example of engineering design considerations being used to derive conclusions about optimal traits; in this case, conclusions about what size prey would optimize net energy intake in mantid prey selection. Dennett's claim that "selection *for*" is intentional rests in part on the fact that such engineering-design considerations guide adaptationist hypotheses about which characteristics were selected *for* in evolutionary history. The idea is that once engineering-design considerations establish which traits are optimal, field investigations can determine whether organisms exhibit those traits. Adaptationists argue that optimal characteristics that are in fact exhibited are plausibly treated as adaptations, that is, characteristics that were selected *for*.

Holling finds that the choice of prey size among mantids is optimal in the sense that engineering-design considerations indicate. Adaptationists invite us to conclude that (the mechanism responsible for) prey-size choice was selected *for*. In this way, optimality arguments connect engineering-design criteria with the biologists' notion of "selection *for*." Such arguments motivate Dennett's claim (*Stance*, p. 318) that there is no middle ground between viewing nature as an engineer and jettisoning the distinction between "selection *of*" and "selection *for*," and his claim (*Précis*, p. 502) that the predictive significance of mentalism in psychology and adaptationism in biology depends on assumptions that are strongly analogous; "the rationality assumption of the intentional stance, . . . and the optimality assumptions of adaptationist thinking." (Similar claims are made by Sober 1981; 1984; 1985.) The main point being made is that practically speaking, optimality (and rationality) principles are essential to the interpretation of evidence concerning adaptation (and intention). It is worth considering whether optimality arguments confirm Dennett's conclusion that explanation in terms of natural selection is like explanation that is mentalistic or teleological in the sense specified by Bennett.

Two main issues need to be addressed: The first is whether optimality arguments are sound. The second is whether optimality principles are teleological. The literature indicates that the first issue is controversial, but at least the issue is clear. As Beatty (1980, p. 545) notes, "an implicit premise in any such attempt to account for optimal design in terms of natural selection is an optimization principle to the effect that natural selection results in the predominance of the optimal forms." The controversy is about whether any such principle can be defended. Informed opinion on this differs. Beatty (1980, p. 545) asserts that "such a principle can be derived from the theory of population genetics." Richard Lewontin apparently demurs. He asserts (1979, p. 12) that "the dynamics of natural selection does not include foresight, and there is no *theoretical* principle that assures optimization as a consequence of selection." This disagreement indicates that the relation between optimization and selection is not as straightforward as Dennett and many adaptationists imply. It is quite clear, however, what endorsing optimality principles entails. Models of natural selection that include such principles make natural selection goal-directed in exactly the sense specified by Bennett. A case of host-parasite interaction (Cowan 1990, p. 201) provides an illustration of this.

Invasion by the castrating trematode (a worm of the genus *Microphallus*) constitutes a critical change in the freshwater habitat of a common New Zealand snail (*Potamopyrgus antipodarum*). In the absence of this parasite, asexual reproduction by the snails yields a higher reproductive rate than sexual reproduction. Invasion by the castrating trematode produces an environment where (long-term) reproductive success increases directly with the likelihood that offspring are resistant to the parasite. Sexual reproduction does increase this likelihood and so from the standpoint of engineering design is optimal (in relation to prevailing conditions) for enhancing fitness. It is also a characteristic that is instrumental in Bennett's sense for increasing reproductive fitness. Optimality principles predict that such optimal traits will emerge, and so the predictive significance of viewing natural selection as an optimization process depends on natural selection's being a process that is teleological in exactly Bennett's sense.

What is problematic here is the relation between optimization principles and the processes of variation and selection that operate in natural selection. Clearly, nonteleological processes of variation and selection do not guarantee optimization, but equally clearly, under certain circumstances optimal traits will be left (e.g., if optimal traits are generated, they will tend to be selected *for*). This poses questions about the process of natural selection and the property of biological fitness. For example, one could "make" natural selection an optimizing process by stipulating that the process of natural selection occurs only under conditions where optimization occurs (i.e., where traits emerge that are fitness enhancing in relation to a specified range of circumstances). Alternatively, one could permit natural selection to occur where something other than optimization (e.g., reproductive disadvantage and an increased likelihood of extinc-

tion) results from critical changes in the environment. To make a long story short, it seems clear that the first alternative is too restrictive and so the status of optimization principles and their relation to selection processes remain problematic. But at least the implication of endorsing them is clear: They make natural selection a process that is goal directed in Bennett's clearly articulated sense. A similar set of points can be made about rationality (e.g., maximization) assumptions in relation to the "generate and test" models of cognitive processes (e.g., rational choice) that Dennett (1978) favors. Dennett needs to consider whether he is prepared to defend a concept of natural selection (and "generate and test" models of cognition) as goal-directed in the sense Bennett articulates.

Dennett need not argue that selection processes (like natural selection or rational choice) necessarily optimize. He can preserve his claims about the links between intentionality and selection *for* by rejecting Bennett's conception of teleology and intention and opting for a conception that can be identified with or explicated in terms of the biologists' notion of selection *for*. So construed, Dennett's views exemplify one type of account of mentalism, function, and teleology that has been presented in the literature. I call it the neo-Darwinian or etiological account. Larry Wright (1976) presents the earliest detailed defense of a view that can incorporate this account, but others (e.g., Bechtel 1986; Brandon 1981; Lehman 1965; Mace 1935; Millikan 1984; 1989; Papineau 1984; Ringen 1976; 1985; Ruse 1971; Wimsatt 1972) have also defended versions of it. My arguments show that this neo-Darwinian (etiological) account is quite different from the one (i.e., Bennett's) that Dennett explicitly endorses. Either Dennett must acknowledge the differences between his views and Bennett's or he must acknowledge that he embraces Bennett's views and distance himself from the neo-Darwinian account; or he must, as he has not yet succeeded in doing, show how distinctive elements of both the neo-Darwinian view and Bennett's view can be consistently combined in explicating the intentional stance. Each of the first two alternatives poses problems for a consistent interpretation of Dennett.

If Dennett were to endorse either the neo-Darwinian (etiological) view or Bennett's view, it would seem necessary for him to disavow many of the things that he suggests are central to the intentional stance. To see why this is the case, consider the hypothesis that the fundamental difference between etiological accounts of function, intention, and teleology and Bennett's account is exactly the difference between explanation in terms of natural selection and explanation in terms of final causation. This is, of course, a historical speculation, but it is a plausible one. Very much in the fashion of Nagel (1977) and Kant (1990), Bennett presents an account of teleology as an intelligible, contentful, and empirically discoverable form of causation that does not involve "backward causation" and is perfectly compatible with the (constitutive but nonreductive materialist) thesis that objects in the world are composed of only material elements. In the spirit of Taylor (1964) and von Wright (1971), Bennett argues that explanations that invoke intentional notions such as intention, belief, and desire, are teleological in the sense he describes. Prima facie, such a view seems very much like Aristotle's view of final causation (see, e.g., Aristotle 1907; Darwin 1923; Driesch 1914; Kant 1990; Kuhn 1962; Lenoir 1982; Ringen 1986), and, as has been shown, quite different from the kind of causation exemplified in natural selection. This contrast has some significance, both for interpreting Dennett's intentions and for interpreting conceptual and methodological practices in contemporary psychology and biology.

Contemporary radical behaviorists (e.g., Skinner 1969; 1974; see also *BBS* special issue, "Canonical Papers of B. F. Skinner" *BBS* 7(4) 1984) and eliminative materialists (e.g., Churchland 1979; Quine 1960; Stich 1983) find the use of mentalistic explanations in cognitive science problematic. Critics of the adaptationist program in biology (e.g., Gould & Lewontin 1979) have serious reservations about the compatibility of optimization principles and many functional explanations with the facts of molecular genetics and natural history. I have indicated why these critics of cognitivism and adaptationism can consistently preserve the distinction between "selection *of* a characteristic" and "selection *for* a characteristic" without importing problematic teleological or mentalistic notions. Similar arguments (see, e.g., Ghiselin 1969) would show how a radical behaviorist can use the principles of operant conditioning to make a similar distinction in psychology. To the extent that Dennett offers an etiological account of function, intention, and teleology, he offers notions that are perfectly compatible with radical behaviorist, eliminative materialist, and antiadaptationist views. This interpretation seems to fit the Dennett (1987) who defends Quine and who wrote "Why the Law of Effect will not go away." This Dennett has made a break with the old teleological (and mentalistic) tradition that is well explicated by Bennett.

There is, of course, a substantial problem with this interpretation that is raised by the specter of another Dennett. This is the Dennett (*Stance*, pp. 283, 286, 347–48) who allies himself with Aristotle and Bennett, and sharply separates himself (*Stance*, pp. 277–82, 348; *Précis*, p. 502) from the antimentalistic and antiteleological views of Skinner, Gould, Lewontin, and the eliminative materialists. This Dennett describes the intentional stance as explicitly mentalistic and teleological and defends cognitivism and adaptationism as instances of the intentional stance. Both this (Aristotelian?) Dennett and the (previously discussed) neo-Darwinian Dennett appear in *The intentional stance* and the *Précis* of it. The relation between these two Dennetts is problematic. The problem can be presented as a dilemma.

If Dennett is endorsing an etiological account of intention, teleology, and function, then there seems to be nothing but a terminological difference between his views and the antimentalist and antiadaptationist views that he explicitly disavows, but there is a significant conceptual difference between his views and those (e.g., Bennett's) that he explicitly endorses. If, on the other hand, Dennett is serious about embracing Bennett's (Aristotelian?) account of teleology and intention, then he seems to be mistaken in thinking that teleology, intention, and function in this sense are inextricably linked to the distinction between selection *for* characteristics and selection *of* them. Furthermore, if my historical hypothesis is correct, then investigation should show that insofar as the intentional stance and the adaptationist stance are linked to teleology and intentionality in Bennett's sense they (unlike natural selection) are linked to the classical (Aristotelian/Kantian) teleological tradition. This would ground Dennett's opposition to Skinner's radical behaviorism, to the antiadaptationism of Gould and Lewontin, and to eliminative materialism, but it would carry with it the implication that according to Dennett's account the core of cognitivist programs in psychology and adaptationist programs in biology is a doctrine that is as old as the Aristotelian doctrine of final causation and is embodied in the developmental vitalism inspired by Kant's (1990) "teleomechanist" conception of biology. In this sense, these views represent a (reactionary?) break from the modern scientific tradition begun by Galileo (Burtt 1951). They suggest that the vitality of modern cognitivism and adaptationism provides evidence that explanation in terms of final causation is an essential (practical) feature of the life sciences. I simply want to ask whether Dennett is uneasy with these consequences of embracing Bennett, and if so, whether he would indicate how to avoid these consequences without being impaled on the other horn of my dilemma. Perhaps this question underscores the significance of Darwin's legacy for naturalized accounts of "purpose" and "intentionality."

ACKNOWLEDGMENTS

The evolution of this essay was facilitated by discussions with Jonathan Bennett, Rob Chametzky, David Depew, David Rider, Terry Smith, and most recently Daniel Dennett. These discussions and, especially, Jon-

athan Bennett's encouraging comments on the first version and Daniel Dennett's constructively critical remarks on a later version are greatly appreciated. The (longer) penultimate version of this essay was read to the Philosophy Faculty Colloquium at the University of Iowa. Much of the writing was done at the center for Advanced Studies at the University of Iowa. The use of the support facilities there is gratefully acknowledged.

Author's Response

Evolution, teleology, intentionality

Daniel C. Dennett
Center for Cognitive Studies, Tufts University, Medford, MA 02155-7068

No response that was not as long and intricate as the two accompanying commentaries combined could do justice to their details, so what follows will satisfy nobody, myself included. I will concentrate on one issue discussed by both commentators: the relationship between evolution and teleological (or intentional) explanation. My response, in its brevity, may have just one virtue: It will confirm some of the hunches (or should I say suspicions) that these and other writers have entertained about my views. For more closely argued defenses of my points, see Dennett (1990a; 1990b; 1990c; 1991a; 1991b).

As **Ringen** notes, I have claimed that mentalistic or intentional explanations are not just similar to adaptationist explanations of evolution but continuous with them; there is just one sort of explanation here, operating according to one set of principles. Ringen thinks this is mistaken, and presents me with a dilemma: I must side either with the neo-Darwinians (who offer to reduce or even eliminate teleology via a mechanistic model of natural selection) or with **Bennett** (who according to Ringen champions a nonreductive, Aristotelian concept of real teleology). The position Bennett presents is more nuanced than Ringen suggests, but he supports a version of Ringen's challenge: He deplores what he sees as my fence-sitting "stance" talk, and urges me to get real. Where I have said that "nothing without a great deal of structural and processing complexity could conceivably realize an intentional system of any interest," Bennett "would replace the last phrase by 'a genuinely intentional system,' leaving 'interest' out of it." He sketches several of his proposals for settling the determinable questions of intentional attribution in ways, he claims, that are independent of my appeals to evolution.

There is a symmetry to **Bennett**'s and **Ringen**'s disagreements with me. Ringen maintains that, contrary to what I have said, the concept of *selection for*, and hence a basis for adaptationist theorizing in biology, can be secured independently of any intentionalizing of the design process – one need not appeal to "what Mother Nature had in mind." Bennett maintains that, contrary to what I have said, assertions about intentional attributions – about what an organism "had in mind" – can be secured independently of any assumptions about the provenance in evolution of the organism in question. If they were both right, we could have a nonintentional evolutionary theory and a nonevolutionary theory of intentionality.

I continue to think they are both wrong. The apparent differences between adaptationist theorizing in biology and intentionalist theorizing in psychology are due, in my view, to the huge differences in time scale, and – more evident in the discussions of both **Ringen** and **Bennett** – a downplaying of the importance of the implications of the ubiquitous idealizing assumptions in both enterprises. When we grasp the nettle and confront the ineliminable "practical difficulties" (Ringen) that beset the evolutionary theorist intent on distinguishing actual cases of *selection for*, and the parallel practical inability of the intentionalist psychologist to cash out the idealizing assumptions that permit talk (in Bennett's example) about a "class of environments . . . unified with help from the concept of food-getting" we see that both enterprises continue to avail themselves – quite appropriately and defensibly – of what Quine called the "dramatic idiom"; the sense-making interpretation-talk of the intentional stance. I claim that since there is just one sort of explanation going on in both quarters, the choice Ringen offers me must be rejected: Teleology is neither as illusory as his neo-Darwinians claim, nor as real and irreducible as his Aristotelian Bennett claims.

Ringen renders usefully explicit the vision of real teleology that haunts current thinking both in evolutionary theory and in philosophy of mind – where I have in mind particularly Dretske's (1986; 1988b) quest for a causal role for meanings. Suppose there were such a thing as a genuinely teleological system, or, equivalently, a *real* (as opposed to approximate or "as-if") intentional system: "Teleological principles provide a basis for predicting what response to new circumstances a system which conforms to them will produce" (Ringen, para. 5; see also Bennett, sect. 7, but note also that Bennett recognizes this to be too idealized, because of the omnipresent possibility of error). Such a system would not just happen to track appropriateness; it would do so in a principled way. It would be *caused*, in Dretske's view, to track meanings in an appropriate way. But there are no such systems, human or otherwise. There are only better and worse approximations of this ideal – which is rather like the ideal of a frictionless bearing, or a perfectly failsafe alarm system. As Ringen points out, the process of natural selection does not quite measure up as a teleological system. Selection itself can only filter, at best supporting the conditional: *If* the appropriate sort of variation is generated, it will be selected. The generation process that provides the candidates for sorting itself is deemed by orthodoxy to be unresponsive to appropriateness. So there can be no guarantee, or anything even close to a guarantee, of genuine "teleological" or meaning-tracking behavior in evolution. I agree, then, with the passage Ringen quotes from Lewontin (1979): "The dynamics of natural selection does not include foresight, and there is no *theoretical* principle that assures optimization as a consequence of selection."

Both **Ringen** and **Bennett** would like to accept the invited contrast of this orthodox view of evolution with a design process controlled by an Intelligent Artificer (or just an intelligent artificer – an everyday, foresightful intentional system such as an engineer). When we look closely at the contrast, however, do we discover anything

but differences in degree? Some engineers are doltish and habit-bound; *if* a particular design solution happens to occur to them, they will adopt it, but there is no guarantee that they will generate the move that we can see in hindsight is the appropriate move in the circumstances. Some engineers are much cleverer, and some have positively brilliant insights into the reasons for and against particular design proposals. How adroit, how flexible, how sensitive must a system be to these reasons for it to be a *real* intentional system?

Bennett's "unity condition" is supposed to answer this question: If "the class of environments is unified with help from the concept of behaving in a manner appropriate" to this or that feature, then we are entitled to attribute that concept to that system, not as a *façon de parler* but literally. But one theorist's unifying concept is another theorist's inflationary shorthand for a mere disjunction of tropisms (cf. Dretske (1986) and Dennett 1987, Ch. 8). Bennett in effect concedes this, for he casts his question in terms of when we may hypothesize that there are going to be more disjuncts than we have observed: "What can make it all right for us to trust an intentional or teleological generalization to lead us from some S-M linkages to predict *others?*" (sect. 7, para. 3). Bennett suggests that either hypotheses about evolution or learning or both could underlie our confidence that one way or another there were mechanisms in an organism (or artifact) that would tend to yield further appropriate linkages. I agree (see Dennett 1990a; 1990b; 1991b), and I do not see (yet) why Bennett claims that his view in this regard is "quite different from" what I have been saying.

I think **Ringen**'s optimism about the independent application of optimality principles in evolutionary theory is similarly undercut. In discussing the case of the sexually reproducing snails' response to the castrating trematode parasite, for instance, he says: "Optimality principles predict that such optimal traits will emerge." I think not. Optimality principles predict that either such optimal traits will emerge or they will not; in the latter case, either the parasites will secure their own extinction by the extinction of their nonadapting hosts, or some semistable exploitation cycle will persist indefinitely. There are no guarantees, only the rationales of hindsight. But do not knock hindsight; one way or another, it is the only sort of sight we can ever count on having. At our best, our adaptive mechanisms lag slightly behind reality, tracking it ever more doggedly, but never giving us a "principle" by which we might predict genuinely teleological activity.

Finally, I will comment all too briefly on some of **Bennett**'s other constructive criticisms and objections. Bennett corrects my interpretation of his views on Quine's 1960 indeterminacy thesis. The view he and Blackburn (1975) hold had not occurred to me, and I have no opinion, yet, on whether, as he claims, determinacy of language is consistent with indeterminacy of thoughts.

Bennett describes my view of intentional attribution as "free-ranging, somewhat haphazard" since it is governed by only "two extremely mild constraints": a rationality assumption and a prohibition of inflationary attributions. He claims that, on the contrary, "a good deal of discipline" can be brought to bear on the project. I have no quarrel with the details of his sketched example (the animal's food-seeking behavior); I just think the considerations he correctly raises are subsumed under my constraints, which are not mild at all. There is plenty of structure to the reasoning processes that govern the postulation and support of intentional attributions, and it is generated, indirectly, by my minimal constraints.

I think Bennett misunderstands the strategy of my vending machine example. He is not alone, so it is my fault. He is right that the vending machine is even worse than the thermostat as an example of an intentional system – that was deliberate on my part. I wanted to choose an example of a dead-simple, quasiperceptual mechanism (the counterfeit-coin detector) so there would be no controversy about "what we would say"; *of course* there is no deep fact of the matter in this instance about which cases of coin-rejection count as "errors." Any grounds for calling some cases errors and others proper functioning will have to depend on the embedding of the device in a large context of purposes: the purposes of its users. The challenge is then for the believers in deeper facts about content in fancier cases (Twin Earth cases in particular) to show what features of these fancier cases permit us to invoke other principles. I claim they cannot, and I do not see that Bennett's discussion provides any such grounds. Bennett says we do not have to solve the problem of error for the vending machine. We do not *have* to solve the problem of error for anything; we can always ("in principle") eschew intentional discourse and settle for brute physical stance mechanism. But *if* we find it illuminating to adopt the intentional stance (and even in the case of the vending machine, the error-talk is illuminating – just think of the design-improvement process, the invocation of Gresham's Law, etc.), we will find ourselves invoking the minimal but none-too-mild constraints of the intentional stance.

References

Amundson, R. (1989) The trials and tribulations of selectionist explanations. In: *Issues in Evolutionary Epistemology*, ed. K. Hahlweg & C. A. Hooker. State University of New York Press. [JR]

Aristotle (1907) *De anima*. (R. B. Hicks, Trans.). Cambridge University Press. [JR]

Beatty, J. (1980) Optimal design models and the strategy of model building in evolutionary biology. *Philosophy of Science* 47:532–61. [JR]

Bechtel, W. (1986) Teleological functional analyses and the hierarchical organization of nature. In: *Current issues in teleology*, ed. N. Rescher. University Press of America. [JR]

Bennett, J. (1976) *Linguistic behaviour*. Cambridge University Press. (2nd edition [1990] published by Hackett.) [JB, JR]

(1983) Cognitive ethology: Theory or poetry? *Behavioral and Brain Sciences* 6:356–58. [JB]

(1991a) How to read minds in behaviour: A suggestion from a philosopher. In: *Natural theories of mind: Evolution, development and simulation of everyday mindreading*, ed. A. Whiten. Basil Blackwell. [JB]

(1991b) How is cognitive ethology possible? In: *Cognitive ethology: The minds of other animals*, ed. C. A. Ristau. Erlbaum. [JB]

(1991c) Folk psychological explanations. In: *The future of folk psychology: Intentionality and cognitive science*, ed. J. D. Greenwood. Oxford University Press. [JB]

Blackburn, S. (1975) The identity of propositions. In: *Meaning, reference and necessity: New studies in semantics*, ed. S. Blackburn. Cambridge University Press. [rDCD, JB]

Block, N. (1980) What is functionalism? In: *Readings in philosophy of psychology*, ed. N. Block. Harvard University Press. [JB]

Brandon, R. (1981) Biological teleology: Questions and explanations. *Studies in History and Philosophy of Science* 12:91–105. [JR]

Burtt, E. (1951) *The metaphysical foundations of modern physical science*. (Revised ed.) Humanities Press. [JR]

Churchland, Paul (1979) *Scientific realism and the plasticity of mind.* Cambridge University Press. [JR]
Cowan, R. (1990) Parasite power. *Science News* 138:200–02. [JR]
Darden, L. & Cain, J. (1989) Selection type theories. *Philosophy of Science* 56:106–29. [JR]
Darwin, C. (1923) *The origin of species by means of natural selection* (Sixth edition). D. Appleton. [JR]
Dennett, D. C. (1974) Why the law of effect will not go away. *Journal of the Theory of Social Behavior* 5:169–87. [JB]
(1978) Why the law of effect will not go away. In: *Brainstorms.* Bradford Books. [JR]
(1987) *The intentional stance.* MIT Press/Bradford Books. [rDCD, JB, JR]
(1988) Précis of *The Intentional Stance* [followed by Open Peer Commentary and Author's Response]. *Behavioral and Brain Sciences* 11:495–546. [JB, JR]
(1990a) Ways of establishing harmony. In: *Dretske and his critics,* ed. B. McLaughlin. Blackwell. [rDCD]
(1990b) The interpretation of texts, people, and other artifacts. *Philosophy and Phenomenological Research* 50:177–94. [rDCD]
(1990c) Dr. Pangloss knows best [reply to Amundson]. *Behavioral and Brain Sciences* 13:581–82. [rDCD]
(1991a) Real Patterns. *Journal of Philosophy* 87:27–51. [rDCD]
(1991b) Do-it-yourself understanding. Center for Cognitive Studies Preprint CSS-90-4, Tufts University, Medford, MA. [rDCD]
Dretske, F. (1985) Machines and the mental. *Proceedings and Addresses of the American Philosophical Association* 59:23–33. [JB]
(1986) *Misrepresentation.* In: *Belief,* ed. R. Bogdan. Oxford University Press. [rDCD, JB]
(1988a) The stance stance. *Behavioral and Brain Sciences* 11(3):511–12. [JB]
(1988b) *Explaining behavior: Reasons in a world of causes.* MIT Press/Bradford Books. [rDCD]
Driesch, H. (1914) *The history and theory of vitalism.* (C. K. Ogden, Trans.) Macmillan. [JR]
Dupré, J. (1987) *The latest on the best: Essays on evolution and optimality.* MIT Press. [JR]
Ghiselin, M. (1969) *The triumph of the Darwinian method.* University of California Press. [JR]
Gould, S. (1980) Caring groups and selfish genes. In: *The Panda's Thumb.* W. W. Norton. (Reprinted in *Conceptual issues in evolutionary biology: An anthology,* ed. E. Sober. MIT Press. [JR]
Gould, S. & Lewontin, R. (1979) The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society.* B205:581–98. Reprinted in *Conceptual issues in evolutionary biology: An anthology,* ed. E. Sober. MIT Press. [JR]
Holling, C. (1964) The analysis of complex population processes. *Canadian Entomologist* 96:335–47. [JR]
Kant, I. (1990) *The critique of practical judgment.* Hackett. [JR]
Kuhn, T. (1962) *The structure of scientific revolutions.* University of Chicago Press. [JR]
Lehman, H. (1965) Functional explanations in biology. *Philosophy of Science* 32:1–20 [JR]
Lenoir, T. (1982) *The strategy of life: Teleology and mechanics in 19th century German biology.* D. Reidel. [JR]
Lewontin, R. (1979) Fitness, survival, and optimality. In: *Analysis of ecological systems,* ed. D. Horn, G. Stairs & R. Mitchell. Ohio State University. [JR]
Mace, C. A. (1935) Mechanical and teleological explanation. In: *Science, History, and Theology* (Aristotelian Society, Supplementary). 14:22–45. Harrison & Sons Ltd. [JR]
Maull, N. (1977) Unifying science without reduction. *Studies in History and Philosophy of Science* 9:143–62. Reprinted in *Conceptual issues in evolutionary biology: An anthology.* MIT Press. [JR]
Millikan, R. (1984) *Language, thought, and other biological categories: New foundations for Realism.* MIT Press. [JR]
(1989) In defense of proper functions. *Philosophy of Science* 56:288–302. [JR]
Nagel, E. (1977) Teleology revisited. *The Journal of Philosophy* 74:61–300.
Papineau, D. (1984) Representation and explanation. *Philosophy of Science.* 51:550–72. [JR]
Quine, W. V. (1960) *Word and object.* MIT Press. [rDCD, JB, JR]
Ringen, J. (1976) Explanation, teleology, and operant behaviorism: A study of the experimental analysis of purposive behavior. *Philosophy of Science* 43:223–53. [JR]
(1985) Operant conditioning and a paradox of teleology. *Philosophy of Science* 52:565–77. [JR]
(1986) The completeness of behavior theory: A review of *Behaviorism, science, and human nature* by Barry Schwartz & Hugh Lacey. *Behaviorism* 14:29–39. [JR]
Ruse, M. (1971) Function statements in biology. *Philosophy of Science* 38:87–95. [JR]
Shanahan, T. (1990) Evolution, phenotypic selection, and the units of selection. *Philosophy of Science* 57:210–25. [JR]
Skinner, B. F. (1969) *Contingencies of reinforcement: A theoretical analysis.* Appleton-Century-Crofts. [JR]
(1984) The evolution of behavior. *Journal of the Experimental Analysis of Behavior* 41:217–21. [JR]
Sober, E. (1981) Evolutionary theory and the ontological status of properties. *Philosophical Studies* 40:147–76. [JR]
(1984) *The nature of selection.* MIT Press. [JR]
(1985) Methodological behaviorism, evolution, and game theory. In: *Sociobiology and epistemology,* ed. J. Fetzer. D. Reidel. [JR]
(1986) Comments on Rosenberg's review. *Behaviorism* 14:89–96. [JR]
Staddon, J. E. R. (1983) *Adaptive behavior and learning.* Cambridge University Press. [JR]
Stich, S. (1983) *From folk psychology to cognitive science: The case against belief.* MIT Press. [JR]
(1988) Connectionism, Realism and realism. *Behavioral and Brain Sciences* 11(3):531–32. [JB]
Taylor, C. (1964) *The explanation of behaviour.* Humanities Press. [JR]
von Wright, G. (1971) *Explanation and understanding.* Cornell University Press. [JR]
Wimsatt, W. (1972) Teleology and the logical structure of function statements. *Studies in the History and Philosophy of Science* 3:1–80. [JR]
Wright, L. (1976) *Teleological explanations.* University of California Press. [JR]