

# Comments on Dennett from a Cautious Ally

Jonathan Bennett

from: *Behavioral and Brain Sciences* 16 (1993), pp. 381–5.

## 1. Indeterminacy of content

In these notes, unadorned page numbers under 350 refer to Dennett (1987)—*The Intentional Stance*, hereafter referred to as *Stance*—and ones over 495 refer to Dennett (1988)—mostly to material by him but occasionally to remarks of his critics. Since the notes 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 doesn’t) that our attributions of mental content can always be highly determinate (pp. 347f). In fact, there are 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 Section 3 below). He seems to represent the process as free-ranging, somewhat haphazard, a matter of guesses and luck which 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 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 Blackburn (1975) and Bennett (1976), 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 towards a thermostat, Dennett says. At noon the thermostat *closed the switch* because it *perceived* that the temperature was below 65 degrees, *wanted* it to be at 65 degrees, 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), sections 21–22.

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 terms of thoughts and wants. **(2)** The next answer is the one I shall highlight shortly. **(3)** We often don’t know the chemical explanation, which entitles us to use intentional one *faute de mieux*. **(4)** 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.

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

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 don’t 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 reason 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 which 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 **(2)**.

(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 confirma-

tion of a much too simple mechanical explanation. . . , this really does disconfirm the fancy intentional level account’ (pp. 542f). It looks as though Dennett is here relying on the unity condition as a mark of the intentional, but he doesn’t ever make this a matter of explicit doctrine as it deserves.

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. Re-reading Dennett (1988), I am struck by how often Dennett’s critics put the question in terms of realism about ‘beliefs and desires’; see for example Dretske and Stich. It’s worth separating two questions. **(i)** How realist should we be about statements of the form ‘x believes that P’ and ‘x wants it to be the case that Q’? **(ii)** 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 as well be explained 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.

However, it would be a poor outlook if those two were our only guides.<sup>1</sup> If we want help in devising mentalistic hypotheses to explain animal behavior, **(1)** doesn’t 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 don’t change, and which is free from practical conflict: its beliefs never bring into play two desires that cannot both be satisfied.<sup>2</sup> 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 towards its mouth.

That plainly invites explanation in terms of simple stimulus-response triggers, giving no purchase to explanation in terms of wants and thoughts. That is clearly implied by the unity condition, which won’t 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 behavior pattern involving a class of environments whose best unified description is that in each of them

the environment is such that there is something the animal can do that will bring it food;

and a class of outputs that are united only in that in each of them

the animal moves in some way that results in its getting food.

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

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

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 section 5 below.)

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

Each of the relevant environments is—given the animal's perceptual apparatus, its 'quality space' etc.—significantly similar to ones 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 **(a)** are relevantly similar to the one in which A occurs and **(b)** 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 can't be replaced by anything unitary that doesn't involve that, and that 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—such that on each relevant occasion the

animal receives a stimulus of that kind, then these behaviors do not support the attribution to it 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 ingesting of food. For more details (of which there are plenty), see Bennett (1991b), pp. 46–8.

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 is always the case that its behavior is apt to get them to think there is a leopard nearby.

(How, in the absence of language, could that be? Well, if the monkeys can use the information that a leopard is nearby in various different 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 by this:

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 the class of environments in the earlier example 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. That will have to pass muster for the animal's wanting to get the others to *believe* there is a leopard nearby: it's 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. It has, incidentally, no tendency 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'.<sup>1</sup> (I am inclined therefore to disagree with what seems to be Ringen's main thesis in his commentary in this issue of *Behavioral and Brain Sciences*.)

He 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 can't 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).

(He 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); 'You are... just a product of natural selection, whose intentionality is thus derivative and *hence* potentially indeterminate' (p. 313). I don't know what warrants 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 that it has to be supplemented; but why?

In his defense of this position, two things are going on.

(i) 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.<sup>2</sup>

Now, there are two standpoints from which one might deny this. (1) *Intentionality conceptually involves an ani-*

<sup>1</sup> Dennett (1987), p. 299. I comment on this similarity in Bennett (1976), pp. 78f, 204–210.

<sup>2</sup> 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)determinateness of mental content keeps cropping up in the chapter; it is allowed to hook into everything, often in ways I don't understand.

*mal's relations to its environment.* Many facts about how the animal did, will, and would move, in what kinds of environment, and with what upshots, 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'll have no quarrel with it here. **(2)** *Any instance of intentionality must be derived from the (near-)intentionality of something else.* Thus, Dennett says, a certain machine can be described as having certain 'desires' and 'beliefs' only on the strength of facts about what it owners want it for; and 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 way of rejecting intrinsic intentionality leads to the thesis I'm now discussing, that animal intentionality depends conceptually 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 in 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

**(ii)** Dennett does present such a defense, pointing to a problem that can arise when we want settle what an animal believes and wants on the basis of its relations to its environments—a problem which 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's pretend that we have here enough of the right sort of complexity to justify interpretation in terms of beliefs and desires (see section 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 that Dennett raises is: What resources do we have to preferring **(a)**? If there can't 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 it is 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.<sup>1</sup>

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 that we are trying to interpret? If it was selected because it generally leads to the ingesting of flies, that gives us a nudge towards **(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—e.g. 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 to 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), Chapters 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 healthiness 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 that, but that 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 oughtn't to be handled in intentional terms at all (see section 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 aren't 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 can't 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,

<sup>1</sup> 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, 'x believes that. . .' is vacuous, does no work, is equivalent to 'It is true that. . .'.

US quarters and Panamanian quarter-balboas; after years of use in the USA 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 seems rather to be (as it were) thinking that it is getting quarter-balboas, and sometimes wrongly thinking this about a US 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 couldn't 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 don't 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 can't 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 don't 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.

That point 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 of 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 non-negotiable 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 doesn't 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, just are explanatory; and attempts to analyse them come to grief if they don't give a central place to that fact.<sup>1</sup>

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), and I shall merely 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. But that link corresponds to a single mechanism, and 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 on 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,

<sup>1</sup> For some examples, see Bennett (1976), section 13.

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 isn't the only solution to the problem that it addresses, as I now show.

**(2)** Secondly 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 employ. 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,<sup>1</sup> 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 didn't 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 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.

---

<sup>1</sup> For reasons given in Dennett (1974).

## Bibliography

**Bennett (1976)**

Jonathan Bennett, *Linguistic Behaviour* (Cambridge University Press). A new edition, with a 20-page Introduction, was published by Hackett in 1990.

**Bennett (1983)**

Jonathan Bennett, 'Cognitive Ethology: Theory or Poetry?', *Behavioral and Brain Sciences* 6, pp. 356–58.

**Bennett (1991a)**

Jonathan Bennett, 'How to Read Minds in Behaviour: a Suggestion from a Philosopher', in A. Whiten (ed.), *Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading* (Basil Blackwell), pp. 97–108.

**Bennett (1991b)**

Jonathan Bennett, 'How is Cognitive Ethology Possible?', in Carolyn A. Ristau (ed), *Cognitive Ethology: the Minds of Other Animals* (Lawrence Erlbaum) pp. 35–49.

**Bennett (1991c)**

Jonathan Bennett, 'Folk Psychological Explanations', in John D. Greenwood (ed.), *The Future of Folk Psychology: Intentionality and Cognitive Science* (Oxford University Press).

**Blackburn (1975)**

Simon Blackburn, 'The Identity of Propositions', in S. Blackburn (ed), *Meaning, Reference and Necessity: New Studies in Semantics* (Cambridge University Press, 1975), pp. 182–205.

**Block (1980)**

Ned Block, 'What is Functionalism?', in N. Block (ed), *Readings in Philosophy of Psychology* (Harvard University Press), pp. 171–184.

**Dennett (1974)**

Daniel C. Dennett, 'Why the Law of Effect Will Not Go Away', *Journal of the Theory of Social Behavior* 5, pp. 169–87.

**Dennett (1987)**

Daniel C. Dennett, *The Intentional Stance* (Bradford Books, 1987)

**Dennett (1988)**

Daniel C. Dennett, Précis of *The Intentional Stance*, followed by Open Peer Commentary and Dennett's responses to that, *Behavioral and Brain Sciences* 11, pp. 495–546.