How is Cognitive Ethology Possible?

Jonathan Bennettt

[From Carolyn A. Ristau (ed), *Cognitive Ethology: the Minds of Other Animals* (Lawrence Erlbaum: Hillsdale, NJ, 1990), pp. 35–49.]

ABSTRACT: Cognitive ethology cannot be done well unless its proximate philosophical underpinnings are got straight; this paper tries to help with that. Cognitive attributions are essentially explanatory—if they did not explain behavior, there would be no justification for them—but it doesn't follow that they explain by providing causes for events that don't have physical causes. To understand how mentalistic attributions do work, we need to focus on the quartet: sensory input, belief, desire, and behavioral output. We also need to be able to study *classes* of sensory inputs—one-shot deals are uninterpretable. The crucial guiding rule is, roughly: The animal's behavior shouldn't be explained by attributing to it the belief that P unless the behavior occurs in sensory circumstances belonging to a class whose members are marked off in some way that involves the concept of P and not in any way that is lower than that. The higher/lower distinction can be understood so that the guiding rule is helpful not only in deciding what thoughts to attribute to an animal but also in deciding whether to attribute any thoughts at all.

1. Introduction

My title asks: By what right can one pass from premises about behavior to conclusions about minds? What ultimately is going on when such inferences are made? An impatient but not unreasonable answer might be the following: these are philosophical questions, which means that we ethologists need not worry about them. We *do* infer mentalistic conclusions from behavioral premises, and there are evidently public standards for doing this, agreed controls governing the inferences, shared bases on which we can rationally debate whether those data support that conclusion, and so on. We can maintain this going concern and get what sensible people will regard as solid results without digging down into its philosophical underlay. Similarly, a physicist can get on with his physics without addressing the problem of the philosophical sceptic: What entitles you to be sure that there is a physical world?

Although it is not unreasonable, that answer is wrong. It is not true that cognitive ethology is being conducted on the basis of shared agreed-upon standards and controls. On the contrary, the field is more tang1ed and disputed than it needs to be because everyday working and arguing standards are insecure and idiosyncratic;. the reason for this is that some underlying philosophical issues have not been properly addressed. I apologize for the dogmatic tone of this statement, but the issue is urgent and important. What is at stake is the integrity of cognitive ethology as a field of intellectual endeavor; and because I do believe in it and think it important, I want to see it equipped with solid enough foundations to support a respectable, coherent, disciplined practice.

That doesn't mean that the foundations need to be explored all the way down. For example, I see no reason why cognitive ethologists have to concern themselves with the issue of mental/material dualism; they can be agnostic about whether inferences from behavioral premises to mentalistic conclusions start inside the physical realm and end outside it. But some fouhdational issues have to be faced.

2. Mind as explanatory

We look at behavior and conjecture that the animal has certain thoughts. Minimally, we conjecture that it wants X and thinks that what it is doing is a way to get X. (l assume without argument that belief and desire—thinking that P and wanting it to be the case that Q—lie at the heart of the cluster of cognitive states that we might attribute to animals.) The first question to be faced is this: In attributing cognitive states to the animal, do we purport to explain its behavior? Yes, for two reasons.

(1) The first is strategic. Our inferences from behavior to cognitive states are intelligible if the conclusions are meant to explain the behavior. For then our procedure falls into the familiar pattern of an inference to the best explanation, like the inference from the fact that the lights went out to the conclusion that a fuse has blown. If, on the other hand, our attributions of mentality don't purport to explain what they are based on, it is hard to see what the 'basing' can consist of.

It might be suggested that what is going on is aesthetic; we look at various specimens of animal behavior and ask ourselves: 'Don't we feel comfortable applying such-and-such cognitive language on the basis of this?' If cognitive ethology were thus merely a matter of inviting one another to respond to the stimuli of animal behavior with the utterance of mentalistic sentences, it would not be a fit activity for competent adults. Anyway, cognitive ethologists clearly don't see their work in that way; on the contrary, they look for conclusions that are supposed to have some chance of being true because they are intellectually supported (not just aesthetically or poetically encouraged) by the behavioral data.

Well, then, perhaps cognitive attributions to animals are just ways of codifying facts about their behavior: ''When an animal behaves thus and so, in such and circumstances, we call that "believing that there is a predator in the undergrowth"'—on an analogy with 'When an animal has such and such behavioral features we call it a "rodent"' or 'When a painting is made thus and so, we call it a "fresco"'. This would reduce the language of cognition to a mere system for classifying facts about behavior. Such a system would be worth having only if mentalistic language could make descriptions of behavior shorter and more compact without loss of content. But clearly the language of cognition does not serve in that role: the required equations, with mentality on one side and behavior on the other, don't exist. And cognitive ethologists don't accord it that role. They think (rightly) that facts about what an animal thinks and wants can help to *explain* how it behaves; and such facts couldn't do that if they were themselves really just facts about how animals behave.

Given those failures, I conclude that what we say about the minds of animals is meant to help explain how they behave.

(2) The second reason for this conclusion is more specific and detailed. The link between thought and behavior essentially requires that thought includes *desire*: No information about what an animal thinks, remembers, concludes, or suspects has the slightest bearing on its behavior except in combination with facts about what it wants. And the concept of desire—or its parent concept, namely goal or purpose—is *essentially* explanatory. Some theorists have not seen this. They have tried to explain how one might arrive at the conclusion that an animal's behavior has G as a goal with this being understood as purely descriptive of the animal, untouched by any suggestion that the animal behaves as it does *because* G is its goal. Such attempts to analyze the concept of desire as purely descriptive and in no way explanatory have all failed so radically as to suggest that the project cannot be carried through because desire: is essentially an explanatory concept. (Nagel, 1979; Tolman, 1932, pp. 10, 13, 21.)

Let us take it, then, that in attributing beliefs and desires to animals we are trying to explain their behavior. What kind of explanation can this be?

The most natural answer is: causal explanation. That was Descartes' view of the matter. He thought we could be entitled to attribute thoughts to others only if their behavior could not have been caused purely by their bodily states; because he thought that the behavior of nonhuman animals could all be physicalistically explained, he was unwilling to credit them with having any thoughts at all.

His contemporary Arnauld (1964–1976, vol. 7) predicted that Descartes would have trouble convincing people that the behavior of other animals could be explained in purely physical terms:

> For at first sight it seems incredible that it can come about, without the assistance of any soul, that the light reflected from the body of a wolf onto the nerves of a sheep should move the minute fibres of the optic nerves, and that on reaching the brain this motion should spread the animal spirits throughout the nerves in the manner necessary to precipitate the sheep's flight. (p. 205)

Descartes certainly wasn't entitled to be dogmatic about this. But nor was it reasonable to be confident that he was wrong, and intuitions of incredibility were worthless—as Spinoza said a few years later—given how little was known about how animals, human and other, are built and how they function. It has been made easier for us than it was for Spinoza to see this, helped as we are by microscopic knowledge of the brain's complexity and by a shift from a mechanical to a chemical and electrical understanding of neural processes.

So if we go Descartes' way, we ought to give up cognitive ethology; the physical causes of animal behavior probably suffice to explain it all, leaving no gaps that have to be filled from outside the physical realm. We don't have to like cognitive ethology to dislike this approach. For one thing, it ties the notion of mentality to a Cartesian dualist understanding of it, according to which mind is something that lies right outside the physical world and causally intrudes into it. Also, it puts the belief that people have thoughts at the mercy of the claim that their behavior cannot be physicalistically explained. Descartes thought that it couldn't, but it would be rash of us to agree with him and foolish to make that agreement our only basis for supposing that people think!

3. Another kind of explanation

I conclude that in attributing beliefs and desires to animals we must be offering noncausal explanations of their behavior. How can this be?

Well, what is needed are fairly reliable generalizations relating beliefs and desires to behavior. The core idea is as follows (Bennett, 1976, chapters 2–4). To say that an animal is behaving with the achievement of G as its goal is to say that it is in a condition C such that: whenever it is in condition C it does whatever it thinks will achieve G. That, though vastly too simple, is the seed crystal from which a complete behavior-based theory of belief and desire can be grown. A crucial fact about it is that it ties the notion of an animal's wanting something or having it as a goal to its falling under some general truth—something to the effect that whenever so-and-so obtains the animal does such-and-such. It is precisely because a generalization must lie in the background of any desire that the attribution of desires can be explanatory:

> 'Why did the animal do A?' 'Because it thought that doing A would achieve G, and it was in a condition C such that whenever it is in condition C it will do whatever it thinks will achieve G.'

Bringing its behavior under that kind of generalization is not causally explaining the behavior. Causal explanations of behavior must always be neurological if materialism is true, and they are probably so even if materialism is false. That is, even if there are mentalistic facts that are entirely additional to anything belonging to the world of matter and things in space, it seems reasonable to suppose that the causes of such facts are always facts about brains. The alternative to this is to suppose that physical causal chains have gaps in them—gaps that are plugged by the intrusion of mental events. That is too much to swallow.

Anyway—and this is the main point—we can hold that mentalistic explanations can be genuinely explanatory and worthwhile without being forced to suppose that they are causal. The reason for this has been well enough expressed in the literature, and I shall merely sketch it here. (Bennett, 1976, section 21; Dennett, 1987, pp. 25–28.)

Let us suppose that every move that an animal makes can be fully causally explained in physiological terms (mostly neural ones). Here is the threat we have to meet:

> A mentalistic explanation of an animal's behavior involves concepts that are superficial and relatively local. Ex hypothesi there is always a properly causal explanation, using concepts that go deeper and spread wider through the physical world—concepts of neurophysiology that ultimately reach down into chemistry and physics. The latter kind of explanation is surely preferable to the former. Granted, we may sometimes have a mentalistic explanation of something an animal does and not be able to explain it physiologically, and that may give explanations of the former kind a weak sort of legitimacy. But that is the best case we can make for explaining behavior mentalistically—namely, that we sometimes have to use those explanations *faute de mieux*—so we ought to be a little embarrassed, apologetic. reticent about them, as we should about anything that we would jettison if we were smarter and more knowledgeable.

That is a good try, but it's not good enough. The fact is that our mentalistic explanations involve groupings that would be missed altogether by neurophysiological explanations. The various things that an animal does because it thinks they would lead to food may have no significant neural common factor or anyway none that they don't share with many episodes that have nothing to do with nutrition; similarly for all its moves aimed at escaping from predators, at getting sexual satisfaction, at caring for its young, at attracting a mate, and so on. Thus, mentalistic explanations of behavior bring out patterns, commonalities, regularities, that. would slip through the net of the most densely informed and theoretically supported neurological explanations. If we are interested in those patterns (Why shouldn't we be'?), that entitles us to be unapologetically interested in the explanations that correspond to them.

Consider a situation in which an animal is threatened by a predator, and its behavior is being predicted by us, in mentalistic terms, and by a superhumanly calculating physiologist who has magical ways of knowing pretty exactly what the animal's detailed brain-states are at this moment. If our cognitive account of the animal is good enough, we may be able to predict that it will climb the nearest tree, say; but if we are less well informed, we may be in a position to predict only that the animal will behave in some way that increases its chances of escaping the predator; we might know that much without knowing any more. For example, we may not be able to rule out the possibility that the animal will lie motionless and silent. Now the superhuman physiologist may be able to predict that the animal will move just exactly thus and so: it will climb that tree in that precise manner every moment of each limb being precisely predicted. But if the physiologist isn't quite that good and can only approximate to a perfect prediction, he won't be able to predict that the animal will improve its chances of escaping the predator. He will be able to predict that the animal will make approximately such-and-such movements, which means that it will go towards the tree and then make climbing-like movements; but for all he can tell, the animal may not quite get to the tree, or may get there and move like a climber but not hold on tight enough, which means that he cannot predict that the animal will do something that is likely to save it from the predator.

The physiologist's inability to predict whether the animal will do anything to improve its chances of escaping matches our inability to say even approximately what kinds of movements the animal will make (e.g. climb, swim, or lie still). Neither basis for prediction is better than the other. They merely cater to different legitimate interests; those interests correspond to different patterns, different classifications of episodes, and that shows up in (among other things) different kinds of *spread of approximation*.

4. The notorious triangle

Behavior shows what an animal thinks only on an assumption about what it wants. Behavior shows what an animal wants only on assumptions about what it thinks. This is the famous belief-desire-behavior triangle. The message that it brings to the cognitive ethologist is: What you must look for to explain your subject's behavior is a cognitive theory that involves both cognitive and conative elements, i.e. that has to do with beliefs and with desires. There is no chance at all of determining one of these first and then moving on to study the other.

But how are we to tackle them both at once? Here is a simple and incorrect recipe for doing so. Attend to a bit of behavior A and think up some belief-desire pair B-D such that if the animal had B-D, it would have done A. Attribute B and D to the animal. Then attend to a second bit of behavior and repeat the process. Continue to do this through all the animal's behavioral history. At the end of this you will have a complete story about what it thought and what it wanted at each moment.

A moment's reflection will show that this libertine procedure is worthless; it is too easy. It allows any given bit of behavior to fall under so many different belief-desire pairs that it isn't interesting or significant to pick arbitrarily on some one of them. The animal climbed the tree because it wanted to get warm and it thought there was a fire up there;

or because it wanted to get food and thought the top of the tree was edible; or because it wanted a sexual partner and thought there was one in the tree; or because. . . One can fabricate such belief-desire explanations at will in perfect conformity with the rule of the libertine procedure; so there is no point in coming up with any of them.

Notice that because the libertine procedure does not connect what the animal thinks or wants at one time with what it thinks or wants at another, the procedure cannot lead us to results that will have predictive value. That is one mark of the fact that these explanations don't explain anything. Another sign of trouble is that the libertine procedure makes no provision for the animal sometimes to believe something falsely.

5. From triangle to square

What the libertine procedure offers us is a glassy surface: because none of our conjectures can run into serious trouble, we are left to skid and slide all over the place with no rough ground on which we can take a stand and no reason to prefer, in a given situation, to attribute one belief-desire pair rather than another. What we need as rough ground is some independent basis for preferring some attributions to others for a particular animal at a particular time.

The way we find this basis is by developing some theory about what the animal is likely to believe when it is in such and such an environment and when its sense organs are in such and such a condition and are oriented in such and such way. This is a theory about the relation between the animal's sensory inputs and its beliefs. When we have that, our theory turns out to have not just the three items in the triangle but four: (a) sensory inputs leading through (b) beliefs and (c) desires to (d) behavioral outputs.

But how are we to get any generalizations about what beliefs the animal is likely to have when it has such and such sensory inputs? To do that, we must see what beliefs are indicated by its behavior, but we have seen that the behavior won't tell us about the animal's beliefs unless we know what it wants. Are we, then, still stuck unless we can get some independent basis for judging the likelihood of various attributions of wants? I think not. All we need is a general assumption to the effect that the animal's desires don't change very quickly. Let's start with the strongest form of that assumption and pretend to be sure that the animal's basic desires are always the same. Then we search for some hypothesis about what those unchanging desires are and some general hypotheses about what the animal is apt to believe in various kinds of environments. We are looking for two bits of theory at once, each under its own constraint (the beliefs must relate systematically to the environments; the desires must always be the same), and together they must satisfy the further constraint of yielding belief-desire attributions that fit the animal's actual behavior.

The idea that the animal's desires are always the same is unrealistic. Let's drop it. Actually, there is no difficulty of principle in allowing that our animal's desires might change slowly over time. That still allows us to proceed in nearly the way I have described, with just a slight weakening of the independent constraint on the attributed desires. But might not some of the desires change quickly. Yes indeed. The animal wanted food an hour ago, but in the interim it has gorged; it wanted to play this morning, but now it is tired; it wanted sex five minutes ago, but now it is recovering from an orgasm. There is no threat to our theory from this kind of desire-change, because changes of this kind are caused by and thus correlated with external observable changes in the animal's condition or circumstances. If our no-change-or-slow-change theory gets bogged down, and can't provide enough true predictions, we can then consider the possibility that the animal undergoes some fast changes of desire, and we can watch for the causes of these changes. If we find them, our control over the theory is restored. And if all goes well. we have a theory that predicts as well as explains. In principle. we can predict what our animal is going to do in the next minute or two because we know (a) what the belief-affecting features of its present environment are and how they affect its beliefs. and (b) what its long-term desires are and what, if any, features of its present circumstances are likely to alter those; so we have a basis for saying what it now thinks and wants, and from this we can infer what it will do.

(What if the animal's basic desires change rapidly with no external indications that this is going on? (The resultant behavior will indicate, too late for prediction, that it *has* gone on.) So far as I can see. the behavior of such an animal would be entirely unpredictable and therefore unexplainable. Such an animal, if there were one, would have to lie outside the purview of the cognitive ethologist.)

That four-point procedure is what saves ethology from the threat of complete and hopeless indeterminacy—the threat that any thought-want attribution can be challenged by some equally well-supported rival because a different belief attribution can always be made safe with help from an appropriate shift in the attributed desire· That is the threat to which we would be open if we followed the libertine procedure; but the anchoring of beliefs not only to behavioral outputs but also to sensory inputs gives us an independent grip, putting gravel under our feet so that we don't skid uncontrollably.

1 don't mean to be offering the assurance of fully determinate results. Daniel Dennett (1987, chapter 2), with help from others, has made an unanswerable case for holding that there could be some indeterminacy: Given two somewhat different accounts of what an animal thinks and believes, there may not always be any fact of the matter as to which is correct. The question of how much determinacy there can be is an empirical one: it's no use pontificating about it before doing the work. I am inclined to agree with Dennett that there is a significant amount, i.e. that in cognitive ethology it is inevitable that the data will underdetermine the theory by a good deal. But I see no reason to think, and neither does Dennett, that the underdetermination goes so far as to subvert the whole endeavor in the manner threatened by a careless and panicky look at the belief-desire-behavior triangle.

6. 'Higher' and 'lower'

There is an impressive amount of disagreement over the conjectures through which ethologists seek to explain animal behavior. All the disagreements seem to have this form: John Doe offers data about animal behavior that he says are best explained by hypothesis H_1 about the animal in question, and Jane Doe says that those data can just as well be explained by hypothesis H_2, which is preferable to H_1 because it attributes less to the animal, is more economical, less generous, in what it says about the animal's mind. I don't have any hard-edged general account of what makes one hypothesis (let us say) 'higher' than another, but some species of the genus are easy enough to mark out.

(1) One hypothesis is higher than another if it attributes cognitive mentality while the other doesn't. For example, H_1 says that the lizard shot out its tongue because it wanted to catch a fly and thought that this was the way to catch one, whereas H_2 says that the lizard shot out its tongue because it received a visual stimulus of kind K and had been habituated—or is hard-wired—to make that kind of tongue movement on the receipt of that type of stimulus.

(2) H_1 is higher than H_2 if the thoughts it attributes are more complex than those attributed by the other. For example, H_1 says 'The dog is digging there because it thinks that it buried a bone there earlier and thinks that buried bones stay put', while H_2 says 'The dog is digging there because it thinks there is a bone there'. We may have evidence that if H_2 is true, H_1 must also be true, i.e. the dog thinks that the bone is there now because it thinks that that's where it put the bone and believes that buried bones stay put. But in the absence of extra reasons for that view of the matter, if the behavioral facts are well enough explained by H_2' then it should win out over H_1'.

(3) If H_2 attributes thoughts that are only about the superficial sensorily given features of things, whereas H_1 attributes thoughts that are not so confined, then H_1 is higher than H_2. For example, H_1 says that the monkey called in that way because it wanted its companions to think there was a leopard nearby; H_2 says that it gave the call because it wanted its companions to climb into trees. Both of these explanations involve cognitive mentality, and neither is clearly much more complex than the other; but H_1 counts as higher because it attributes to the animal a desire to make the others believe something (a thought about a thought) whereas H_2 attributes merely a desire to get the others to do something (a thought about a movement).

I have a few ideas about how to pull all this together into a unitary generic account of what it is for one hypothesis to be higher than another, but they are still too incomplete to be worth presenting. Supposing (as I now shall) that we have a usable notion of higher and lower, and that we agree that we ought always to prefer the lowest hypothesis that will satisfactorily explain the behavioral facts, what use are we to make of this in practice?

7. Testing mentalistic hypotheses

Dennett (1983) offered some help with this in his first venture into real-world cognitive ethology. In essence, his offering had two items in it: (a) the principle of rationality, which says that *ceteris paribus* we are entitled to assume that an animal will do what it thinks will achieve its goal, and (b) Lloyd Morgan's canon, which says that *ceteris paribus* we should always prefer the lower to the higher of two mentalistic hypotheses.

The former of these is right. Indeed, it is fundamental to the project of cognitive ethology in a way that Dennett does not bring out. The most elemental move that gets cognitive ethology under way is that of principles about what the animal is likely to believe in given kinds of environments and finding goals that can be attributed to it, that will let one reasonably conjecture that the animal does A because it wants G and thinks that doing A will produce G. Without that last part, which is just an application of the rationality principle, neither beliefs nor desires can be connected with behavior at all.

Dennett's other offering is right too. Without it, cognitive ethology is possible but is so unconstrained, so undisciplined, as not to be worth doing.

The two offerings jointly constitute a testing procedure for mentalistic hypotheses. To test hypothesis H, according to which the animal thinks that P and wants G, attend to it in a situation where that thought and that belief would lead a rational animal to do A and see whether it does A.

If it doesn't, H is false. If it survives that test, there is another to which it can be subjected. Examine the animal's behavior that might be explained by H, and consider whether some lower hypothesis might explain it just as well. If so, then H is condemned as unacceptable. In my not very satisfactory comments on that paper of Dennett's, I was (without realizing it) struggling with two thoughts at once: (a) Dennett had provided only a negative testing procedure, and cognitive ethologists need help with devising hypotheses in the first place, and (b) Dennett hadn't said enough about what it is for a given hypothesis to fit or cover or prima facie explain a range of behavioral data. He left to mere intuition the decisions about what a given hypothesis has in the way of prima facie rivals. I don't now think that there is much in the first of these two ideas, but there is some force in the second. It would be a pity if we couldn't get beyond Dennett's two-part recipe, so that when an ethologist came up with a mentalistic hypothesis and tried and failed to find any lower rivals to it, he merely sat trembling, hoping that no more ingenious and mean-minded colleague would succeed where he had failed. Could we not at least provide some general guide concerning where and how to look for rivals? I think we can, as I now try to show.

8. The guiding rule

For simplicity's sake, let us suppose that our animal's goals don't change, that each of them generates desires that the animal sometimes thinks it can satisfy through its own behavior, and that such beliefs never bring into play two desires that cannot both be satisfied (Bennett. 1976, sections 18–20). Then all we have to look at are the different situations in which this animal thinks there is something it can do that will lead to the satisfaction of one of its desires. Let us look into the class of behavioral episodes in which we think that the animal aims to satisfy desire D, for example, the desire 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. Why? For two reasons: (a) The class of situations in which the behavior occurs can be marked out without reference to anything of the form 'evidence that doing A will produce food', and (b) the class of behaviors can be marked out without reference to 'getting food'. The facts are adequately caught in the statement that whenever the animal has such and such a stimulus-kind of input, it produces such and such a motor-kind of output.

Here is a first approximation to the contrasting case: The class of behaviors to be generalized over involves inputs whose simplest or only 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 involves outputs that are united only in that in each of them the animal moves in some way that results in its getting food. But that is only a first approximation. It would be right only if our animal never went wrong about what would bring it food. I am content to use simplifying, idealizing assumptions so that the discussion doesn't get bogged down in details, but the possibility of error is too important to be idealized away.

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 and 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 bring it food if the environment were a member of the comparison set. Of course in most cases the environment is a member of the comparison set, but if we don't make allowances for the possibility of error, our account will be too drastically idealized and oversimplified.

(There might be some slight misperformance on the animal's part—a slip of the paw, a tiny but significant failure—of such a sort that even if the environment were as the animal thinks it is, the goal still wouldn't be achieved. But we would want such a behavioral episode to be explainable in terms of the animal's having that goal. So, strictly speaking, I ought to have said '. . . moves in a way that would be likely to bring it food if. . . ' or '. . . moves in a way that would nearly bring it food if. . . '. That would provide for the possibility of failure of execution, which is different from failure as a result of cognitive error.)

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 it 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 received a stimulus of that kind, then these behaviors would not 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.*

My guiding rule applies not only to the question of whether it is all right to attribute content, but also to the question of what content to attribute. Did the monkey want its companions to *believe there was a leopard nearby* or merely to *climb a tree*? (I am assuming that the former is higher than the latter.) To have decent evidence that the former attribution is right, we need 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. If the monkeys can use the information that a leopard is nearby in various ways, and animal X's warning cries occur when any one of these uses could be made of the information, the relevant class of environments is marked off as containing all and only environments where X 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, which justifies putting food-getting into the animal's goal and thus its belief, so in this example 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 goal and into its belief. And that is going to pass muster for the animal's having a goal and a belief concerning the others' *believing* that there is a leopard nearby: in the absence of language, there is no chance of getting nearer than that to thoughts and wants regarding the beliefs of others.

9. The output side of the story

What about the output side? I said that we get from a stimulus-response explanation of the behavior to a cognitive one through the move from (1_I) the case where the relevant inputs belong to a single stimulus-kind to (2_I) the case where they are united only by their similarity to situations where the animal can get food. Could we not also make the ascent from stimulus-response to cognition through a move from (1_O) the case where the relevant outputs belong to a single motor-kind to (2_O) the case where they are united only by their being (to put it briefly) apt for the getting of food? What about the output side? I said that we get from a stimulus-response explanation of the behavior to a cognitive one through the move from (1_I) the case where the relevant inputs belong to a single stimulus-kind to (2_I) the case where they are united only by their similarity to situations where the animal can get food. Could we not also make the ascent from stimulus-response to cognition through a move from (1_O) the case where the relevant outputs belong to a single motor-kind to (2_O) the case where they are united only by their being (to put it briefly) apt for the getting of food?

The move from (1_O) to (2_O) is certainly not needed. If the inputs have the right kind of unity, it doesn't matter if the outputs have a unity of a lower kind. Suppose that in each member of the class of episodes we are interested in, the animal simply utters a warning call—there is no significant variation from call to call, but there is a great variation in the physical kinds of situation in which the call is uttered because the animal takes a wide variety of different states of affairs to be clues to the presence of a predator. There is good enough reason here to say that the animal's warning calls are evidence that it thinks there is a predator nearby even though the relevant complexity is all on the input side, with none in the output. (1 here modify slightly the stand I took in Bennett 1976, in the light of a criticism in Peacocke, 1981, p. 216.) I presented just such a case in section 8.

Then is the move from (1_O) to (2_O) sufficient on its own to justify the attribution of cognitive content? There is no answer to that question because there couldn't be such a case, that is, one where there is a lower unity in the inputs but only a higher unity in the outputs. If there were, the animal's pursuits of a certain kind of goal would be triggered by some relatively simple kind of stimulus with no significant differences among the occasions on the input side, but would be executed by a variety of different kinds of movements that have in common only their being apt to produce the goal. For example, a certain characteristic kind of smell or sound sometimes leads the animal to run, sometimes to climb, sometimes to dig, and usually the behavior in question leads to its getting food. This is just magic. In the actual unmagical world, appropriate behavioral variation is made possible by matching variation of sensory clues: The animal jumps to the left one time and to the right next time because of differences in what it sees or hears, or smells or feels. But here we have a story that credits the animal with a useful behavioral variation while excluding any possible explanation of how that is managed.

References

Arnauld, A. (1964–1976). Fourth objections to Descartes's *Meditations*. In C. Adam and P. Tannery (eds.), *Oeuvres de Descartes*, vol. 7, Vrin: Paris, p. 205.

Bennett, J. (1976). *Linguistic Behaviour*. Cambridge University Press.

Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA: M.I.T. Press.

Dennett, D. C. (1983). 'Intentional systems in cognitive ethology: the "Panglossian Paradigm" defended'. *Behavioral and Brain Sciences* 6, 343–90.

Nagel, E. (1979). 'Teleology revisited', in *Teleology and other essays* (pp. 275–316). New York: Columbia University Press.

Peacocke, C. (1981). 'Demonstrative thought and psychological explanation', *Synthese* 49, 187–217.

Tolman, E. C. (1932). *Purposive Behavior in Men and Animals*. New York: The Century Co., 1932.