THE BEHAVIORAL AND BRAIN SCIENCES (1983) 6, 343–390
Printed in the United States of America

Intentional systems in cognitive ethology: The "Panglossian paradigm" defended

Daniel C. Dennett
Department of Philosophy, Tufts University, Medford, Mass. 02155

Abstract: Ethologists and others studying animal behavior in a "cognitive" spirit are in need of a descriptive language and method that are neither anachronistically bound by behaviorist scruples nor prematurely committed to *particular* "information-processing models." Just such an interim descriptive method can be found in *intentional system theory*. The use of intentional system theory is illustrated with the case of the apparently communicative behavior of vervet monkeys. A way of using the theory to generate data – including usable, testable "anecdotal" data – is sketched. The underlying assumptions of this approach can be seen to ally it directly with "adaptationist" theorizing in evolutionary biology, which has recently come under attack from Stephen Gould and Richard Lewontin, who castigate it as the "Panglossian paradigm." Their arguments, which are strongly analogous to B. F. Skinner's arguments against "mentalism," point to certain pitfalls that attend the careless exercise of such "Panglossian" thinking (and rival varieties of thinking as well), but do not constitute a fundamental objection to either adaptationist theorizing or its cousin, intentional system theory.

Keywords: adaptation; animal cognition; behaviorism; cognitive ethology; communication; comparative psychology; consciousness; evolution; intentionality; language; mind; sociobiology

I. The problem

The field of cognitive ethology provides a rich source of material for the philosophical analysis of meaning and mentality, and even holds out some tempting prospects for philosophers to contribute fairly directly to the development of the concepts and methods of another field. As a philosopher, an outsider with only a cursory introduction to the field of ethology, I find that the new ethologists, having cast off the straightjacket of behaviorism and kicked off its weighted overshoes, are looking about somewhat insecurely for something presentable to wear: They are seeking a theoretical vocabulary that is powerfully descriptive of the data they are uncovering and at the same time a theoretically fruitful method of framing hypotheses that will *eventually* lead to information-processing models of the nervous systems of the creatures they are studying (see Roitblat 1982). It is a long way from the observation of the behavior of, say, primates in the wild to the validation of neurophysiological models of their brain activity, and finding a sound interim way of speaking is not a trivial task. Since the methodological and conceptual problems confronting the ethologists appear to me to bear striking resemblances to problems I and other philosophers have been grappling with recently, I am tempted to butt in and offer, first, a swift analysis of the problem, second, a proposal for dealing with it (which I call intentional system theory), third, an analysis of the continuity of intentional system theory with the theoretical strategy or attitude in evolutionary theory often called *adaptationism,* and finally, a limited defense of adaptationism (and its cousin, intentional system theory) against recent criticisms by Stephen J. Gould and Richard C. Lewontin.

The methodology of philosophy, such as it is, includes as one of its most popular (and often genuinely fruitful) strategies the description and examination of entirely imaginary situations, elaborate thought experiments that isolate for scrutiny the presumably critical features in some conceptual domain. In *Word and Object,* W. V. O. Quine (1960), gave us an extended examination of the evidential and theoretical tasks facing the "radical translator," the imaginary anthropologist-linguist who walks into an entirely alien community – with no string of interpreters or bilingual guides – and who must figure out, using whatever scientific methods are available, the language of the natives. Out of this thought experiment came Quine's thesis of the "indeterminacy of radical translation," the claim that it must always be possible in principle to produce nontrivially different translation manuals, equally well supported by all the evidence, for any language. One of the most controversial features of Quine's position over the years has been his uncompromisingly behaviorist scruples about how to characterize the task facing the radical translator. What happens to the task of radical translation when you give up the commitment to a behavioristic outlook and terminology? What are the prospects for fixing on a unique translation of a language (or a unique interpretation of the "mental states" of a being) if one permits oneself the vocabulary and methods of "cognitivism"? The question could be explored via other thought experiments, and has been in some regards (Bennett 1976; Dennett 1971; Lewis 1974), but the real-world researches of Seyfarth, Cheney, and

 0140-525X/83/030343-48/$06.00

Marler (1980) with vervet monkeys in Africa will serve us better on this occasion. Vervet monkeys form societies, of sorts, and have a language, of sorts, and of course there are no bilingual interpreters to give a boost to the radical translators of Vervetese. This is what they find:

> Vervet monkeys give different alarm calls to different predators. Recordings of the alarms played back when predators were absent caused the monkeys to run into the trees for leopard alarms, look up for eagle alarms, and look down for snake alarms. Adults call primarily to leopards, martial eagles, and pythons, but infants give leopard alarms to various mammals, eagle alarms to many birds, and snake alarms to various snakelike objects. Predator classification improves with age and experience. (Abstract of Seyfarth, Cheney & Marler 1980, p. 801)

This abstract is couched, you will note, in almost pure Behaviorese – the language of Science even if it is no longer exclusively the language of science. It is just informative enough to be tantalizing. How much of a language, one wants to know, do the vervets really have? Do they *really* communicate? Do they *mean what they say?* Just what interpretation can we put on these activities? What, if anything, do these data tell us about the cognitive capacities of vervet monkeys? In what ways are they – must they be – like human cognitive capacities, and in what ways and to what degree are vervets more intelligent than other species by virtue of these "linguistic" talents? These loaded questions – the most natural ones to ask under the circumstances – do not fall squarely within the domain of any science, but whether or not they are the right questions for the scientist to ask, they are surely the questions that we all, as fascinated human beings learning of this apparent similarity of the vervets to us, want answered.

The cognitivist would like to succumb to the temptation to use ordinary mentalistic language more or less at face value, and to respond directly to such questions as: What do the monkeys *know*? What do they *want,* and *understand,* and *mean*? At the same time, the primary point of the cognitivists' research is not to satisfy the layman's curiosity about the relative IQ, as it were, of his simian cousins, but to chart the cognitive *talents* of these animals on the way to charting the cognitive *processes* that explain those talents. Could the everyday language of belief, desire, expectation, recognition, understanding, and the like also serve as the suitably rigorous abstract language in which to describe cognitive competences?

I will argue that the answer is yes. Yes, if we are careful about what we are doing and saying when we use ordinary words like "believe" and "want," and if we understand the assumptions and implications of the strategy we must adopt when we use these words.

The decision to conduct one's science in terms of beliefs, desires, and other "mentalistic" notions, the decision to adopt "the intentional stance," as I call it (Dennett 1971; 1976; 1978a; 1981a; 1981b; 1981c), is not an unusual sort of decision in science. The basic strategy of which this is a special case is familiar: changing levels of explanation and description in order to gain access to greater predictive power or generality – purchased, typically, at the cost of submerging detail and courting trivialization on the one hand and easy falsification on the other. When biologists studying some species choose to call something in that species' environment *food* and leave it at that, they ignore the tricky details of the chemistry and physics of nutrition, the biology of mastication, digestion, excretion, and the rest. Even supposing many of the details of this finer-grained biology are still ill-understood, the decision to leap ahead, in anticipation of fine-grained biology, and rely on the well-behavedness of the concept of food at the level of the theory appropriate to it, is likely to meet approval from the most conservative risk takers.

The decision to adopt the intentional stance is riskier. It banks on the soundness of some as yet imprecisely described concept of information – not the concept legitimized by Shannon-Weaver information theory (Shannon 1949), but rather the concept of what is often called *semantic information.* (A more or less standard way of introducing the still imperfectly understood distinction between these two concepts of information is to say that Shannon-Weaver theory measures the *capacity* of information-transmission and information-storage vehicles, but is mute about the *contents* of those channels and vehicles, which will be the topic of the still-to-be-formulated theory of semantic information. See Dretske 1981 [and multiple book review in *BBS* 6(1) 1983] for an attempt to bridge the gap between the two concepts.) Information, in the semantic view, is a perfectly real but very abstract commodity, the storage, transmission, and transformation of which is informally – but quite surefootedly – recounted in ordinary talk in terms of beliefs and desires and the other states and acts philosophers call *intentional.*

II. Intentional system theory

Intentionality, in philosophical jargon, is – in a word – *aboutness.* Some of the things, states, and events in the world have the interesting property of *being about* other things, states, and events; figuratively, they point to other things. This arrow of reference or aboutness has been subjected to intense philosophical scrutiny and has engendered much controversy. For our purposes, we can gingerly pluck two points from this boiling cauldron, oversimplifying them and ignoring important issues tangential to our concerns.

First, we can mark the presence of intentionality – aboutness – as the topic of our discussions by marking the presence of a peculiar *logical* feature of all such discussion. Sentences attributing intentional states or events to systems use idioms that exhibit *referential opacity:* they introduce clauses in which the normal, permissive, substitution rule does not hold: This rule is simply the logical codification of the maxim that a rose by any other name would smell as sweet. If you have a true sentence, so runs the rule, and you alter it by replacing a term in it by another, different term that still refers to exactly the same thing or things, the new sentence will also be true. Ditto for false sentences – merely changing the means of picking out the objects the sentence is about cannot turn a falsehood into a truth. For instance, suppose Bill is the oldest kid in class; then if it is true that

1. Mary is sitting next to Bill,

then, substituting "the oldest kid in class" for "Bill," we get

2. Mary is sitting next to the oldest kid in class, which *must* be true if the other sentence is.

A sentence with an *intentional idiom* in it, however, contains a clause in which such substitution can turn truth into falsehood and vice versa. (This phenomenon is called *referential opacity* because the terms in such clauses are shielded or insulated by a barrier to logical analysis, which normally "sees through" the terms to the world the terms are about.) For example, Sir Walter Scott wrote *Waverly*, and Bertrand Russell (1905) assures us

3. George IV wondered whether Scott was the author of *Waverly*,

but it seems unlikely indeed that

4. George IV wondered whether Scott was Scott.

(As Russell remarks, "An interest in the law of identity can hardly be attributed to the first gentleman of Europe; 1905, p. 485.) To give another example, suppose we decide it is true that

5. Burgess fears that the creature rustling in the bush is a python

and suppose that in fact the creature in the bush is Robert Seyfarth. We will not want to draw the conclusion that

6. Burgess fears that Robert Seyfarth is a python.

Well, in one sense we do, you say, and in one sense we also want to insist that, oddly enough, King George *was* wondering whether Scott was Scott. But that's not how he put it to himself – and that's not how Burgess conceived of the creature in the bush, either – that is, *as* Seyfarth. It's the sense of conceiving *as*, seeing *as*, thinking of *as* that the intentional idioms focus on.

One more example: Suppose you think your next-door neighbor would make someone a good husband and suppose, unbeknownst to you, he's the Mad Strangler. Although in one, very strained, sense you could be said to believe that the Mad Strangler would make someone a good husband, in another more natural sense you don't, for there is another – very bizarre and unlikely – belief that you surely don't have which could better be called the belief that the Mad Strangler would make a good husband.

It is this resistance to substitution, the insistence that for *some* purposes how you call a rose a rose makes all the difference, that makes the intentional idioms ideally suited for talking about the ways in which information is represented in the heads of people – and other animals.

So the first point about intentionality is just that we can rely on a marked set of idioms to have this special feature of being sensitive to the *means of reference* used in the clauses they introduce. The most familiar such idioms are "believes that," "knows that," "expects (that)," "wants (it to be the case that)," "recognizes (that)," "understands (that)."

In short, the "mentalistic" vocabulary shunned by behaviorists and celebrated by cognitivists is quite well picked out by the logical test for referential opacity.

The second point to pluck from the cauldron is somewhat controversial, although it has many adherents who have arrived at roughly the same conclusion by various routes: the use of intentional idioms carries a presupposition or assumption of *rationality* in the creature or system to which the intentional states are attributed. What this amounts to will become clearer if we now turn to the intentional stance in relation to the vervet monkeys.

III. Vervet monkeys as intentional systems

To adopt the intentional stance toward these monkeys is to decide – tentatively, of course – to attempt to characterize, predict, and explain their behavior by using intentional idioms, such as "believes" and "wants," a practice that assumes or presupposes the rationality of the vervets. A vervet monkey is, we will say, an *intentional system*, a thing whose behavior is predictable by attributing beliefs and desires (and, of course, rationality) to it. *Which* beliefs and desires? Here there are many hypotheses available, and they are testable in virtue of the rationality requirement. First, let us note that there are different grades of intentional systems.

A *first-order* intentional system has beliefs and desires (etc.) but no beliefs and desires *about* beliefs and desires. Thus all the attributions we make to a merely first-order intentional system have the logical form of

7. x *believes that* p

8. y *wants that* q

where "p" and "q" are clauses that themselves contain no intentional idioms. A *second-order* intentional system is more sophisticated; it has beliefs and desires (and no doubt other intentional states) about beliefs and desires (and other intentional states) – both those of others and its own. For instance

9. x *wants* y to *believe* that x is hungry

10. x *believes* y *expects* x to jump left

11. x *fears* that y *will discover* that x has a food cache

A *third-order intentional system is one that is capable of such states as*

12. x *wants* y to *believe* that x *believes* he is all alone

A *fourth-order* system might *want* you to *think* it *understood* you to be *requesting* that it leave. How high can we human beings go? "In principle," forever, no doubt, but in fact I suspect that you wonder whether I realize how hard it is for you to be sure that you understand whether I mean to be saying that you can recognize that I can believe you to want me to explain that most of us can keep track of only about five or six orders, under the best of circumstances. See Cargile (1970) for an elegant but sober exploration of this phenomenon.

How good are vervet monkeys? Are they really capable of third-order or higher-order intentionality? The question is interesting on several fronts. First, these orders ascend what is *intuitively* a scale of intelligence; higher-order attributions strike us as much more sophisticated, much more human, requiring much more intelligence. There are some plausible diagnoses of this intuition. Grice (1957, 1969) and other philosophers (see especially Bennett 1976) have developed an elaborate and painstakingly argued case for the view that genuine *communication*, speech acts in the strong, human sense of the word, depend on *at least* three orders of intentionality in both speaker and audience.

Not all interactions between organisms are communicative. When I swat a fly I am not communicating with it, nor am I if I open the window to let it fly away. Does a sheep dog, though, communicate with the sheep it herds? Does a beaver communicate by slapping its tail, and do bees communicate by doing their famous dances? Do human infants communicate with their parents? At what point can one be sure one is really communicating with an infant? The presence of specific linguistic tokens seems neither sufficient nor necessary. (I can use English commands to get my dog to do things, but that is at best a pale form of communication compared to the mere raised eyebrow by which I can let someone know he should change the topic of our conversation.) Grice's theory provides a better framework for answering these questions. It defines intuitively plausible and formally powerful criteria for communication that involve, at a minimum, the correct attribution to communicators of such third-order intentional states as

13. Utterer *intends* Audience to *recognize* that Utterer *intends* Audience to produce response *r*

So one reason for being interested in the intentional interpretation of the vervets is that it promises to answer – or at least help answer – the questions: Is this behavior really linguistic? Are they really communicating? Another reason is that higher-orderedness is a conspicuous mark of the attributions speculated about in the sociobiological literature about such interactive traits as reciprocal altruism. It has even been speculated (by Trivers 1971), that the increasing complexity of mental representation required for the maintenance of systems of reciprocal altruism (and other complex social relations) led, in evolution, to a sort of brain-power arms race. Humphrey (1976) arrives at similar conclusions by a different and in some regards less speculative route. There may then be a number of routes to the conclusion that higher-orderedness of intentional characterization is a deep mark – and not just a reliable symptom – of intelligence.

(I do not mean to suggest that these orders provide a uniform scale of any sort. As several critics have remarked to me, the first iteration – to a *second-order* intentional system – is the crucial step of the recursion; once one has the principle of *embedding* in one's repertoire, the complexity of what one can then in some sense entertain seems plausibly more a limitation of memory, or attention span, or "cognitive workspace" than a fundamental measure of system sophistication. And thanks to "chunking" and other, artificial, aids to memory, there seems to be no *interesting* difference between, say, a fourth-order and a fifth-order intentional system. But see Cargile 1970 for further reflections on the natural limits of iteration.)

But now, back to the empirical question of how good the vervet monkeys are. For simplicity's sake, we can restrict our attention to a single apparently communicative act by a particular vervet, Tom, who, let us suppose, gives a leopard alarm call in the presence of another vervet, Sam. We can now compose a set of competing intentional interpretations of this behavior, ordered from high to low, from romantic to killjoy. Here is a (relatively) romantic hypothesis (with some variations to test in the final clause):

4th-order: Tom *wants* Sam to *recognize* that Tom *wants* Sam to *believe* that there is a leopard
there is a carnivore
there is a four-legged animal
there is a live animal bigger than a breadbox

A less exciting hypothesis to confirm would be this third-order version (there could be others):

3rd-order: Tom *wants* Sam to *believe* that Tom *wants* Sam to run into the trees.

Note that this particular third-order case differs from the fourth-order case in changing the speech act category: on this reading the leopard call is an imperative (a request or command) not a declarative (informing Sam of the leopard). The important difference between imperative and declarative interpretations (see Bennett 1976, §§ 41, 51) of utterances can be captured – and then telltale behavioral differences can be explored – at any level of description above the second order – at which, *ex hypothesi*, there is no intention to utter a speech act of either variety. Even at the second order, however, a related distinction in effect-desired-in-the-Audience is expressed, and is in principle behaviorally detectable, in the following variations:

2nd-order: Tom *wants* Sam to *believe*
that there is a leopard
he should run into the trees

This differs from the previous two in not supposing Tom's act involves ("in Tom's mind") any recognition by Sam of his (Tom's) own role in the situation. If Tom could accomplish his end equally well by growling like a leopard, or just somehow attracting Sam's attention to the leopard without Sam's recognizing Tom's intervention, this would be only a second-order case. (Cf. I *want* you to *believe* I am not in my office; so I sit very quietly and don't answer your knock. That is not communicating.)

1st-order: Tom *wants* to cause Sam to run into the trees (and he has this noise-making trick that produces that effect; he uses the trick to induce a certain response in Sam).

On this reading the leopard cry belongs in the same general category with coming up behind someone and saying "Boo!" Not only does its intended effect not depend on the victim's recognition of the perpetrator's intention; the perpetrator does not need to have any conception at all of the victim's mind: Making loud noises behind certain things just makes them jump.

0-order: Tom (like other vervet monkeys) is prone to three flavors of anxiety or arousal: leopard anxiety, eagle anxiety, and snake anxiety.[1] Each has its characteristic symptomatic vocalization. The effects on others of these vocalizations have a happy trend, but it is all just tropism, in both utterer and audience.

We have reached the killjoy bottom of the barrel: an account that attributes no mentality, no intelligence, no communication, no intentionality at all to the vervet. Other accounts at the various levels are possible, and some may be more plausible; I chose these candidates for simplicity and vividness. Lloyd Morgan's canon of parsimony enjoins us to settle on the most killjoy, least romantic hypothesis that will account systematically for the observed and observable behavior, and for a long time the behaviorist creed that the curves could be made to fit the data well at the lowest level prevented the exploration of the case that can be made for higher-order, higher-level systematizations of the behavior of such animals. The claim that *in principle* a lowest-order story can always be told of any animal behavior (an entirely physiological story, or even an abstemiously behavioristic story of unimaginable complexity) is no longer interesting. It is like claiming that in principle the concept of food can be ignored by biologists – or the concept of cell or gene for that matter – or like claiming that in principle a purely electronic-level story can be told of any computer behavior. Today we are interested in asking what gains in perspicuity, in predictive power, in generalization, might accrue if we adopt a higher-level hypothesis that takes a risky step into intentional characterization.

The question is empirical. The tactic of adopting the intentional stance is not a matter of *replacing* empirical investigations with aprioristic ("armchair") investigations, but of using the stance to suggest which brute empirical questions to put to nature. We can test the competing hypothesis by exploiting the rationality assumption of the intentional stance. We can start at either end of the spectrum; either casting about for the depressing sorts of evidence that will *demote* a creature from a high-order interpretation, or hunting for the delighting sorts of evidence that *promote* creatures to higher-order interpretations (cf. Bennett 1976). We are delighted to learn, for instance, that lone male vervet monkeys, traveling between bands (and hence out of the hearing – so far as they know – of other vervets) will, on seeing a leopard, *silently* seek refuge in the trees. So much for the killjoy hypothesis about leopard-anxiety yelps. (No hypothesis succumbs quite so easily, of course. Ad hoc modifications can save any hypothesis, and it is an easy matter to dream up some simple "context" switches for leopard-anxiety yelp mechanisms to save the zero-order hypothesis for another day.) At the other end of the spectrum, the mere fact that vervet monkeys apparently have so few different things they can say holds out little prospect for discovering any real theoretical utility for such a fancy hypothesis as our fourth-order candidate. It is only in contexts or societies in which one must rule out (or in) such possibilities as irony, metaphor, storytelling, and illustration ("second-intention" uses of words, as philosophers would say)[2] that we must avail ourselves of such high-powered interpretations. The evidence is not yet in, but one would have to be romantic indeed to have high expectations here. Still, there are encouraging anecdotes.

Seyfarth reports (in conversation) an incident in which one band of vervets was losing ground in a territorial skirmish with another band. One of the losing-side monkeys, temporarily out of the fray, seemed to get a bright idea: it suddenly issued a leopard alarm (in the absence of any leopards), leading *all* the vervets to take up the cry and head for the trees – creating a truce and regaining the ground his side had been losing. The intuitive sense we all have that this is *possibly* (barring killjoy reinterpretation) an incident of great cleverness is amenable to a detailed diagnosis in terms of intentional systems. If this act is not just a lucky coincidence, then the act is truly devious, for it is not simply a case of the vervet uttering an *imperative* "get into the trees" in the expectation that *all* the vervets will obey, since the vervet (being rational – our predictive lever) should not *expect* a rival band to honor *his* imperative. So either the leopard call is *considered* by the vervets to be informative – a *warning*, not a *command* – and hence the utterer's credibility but not authority is enough to explain the effect, or our utterer is more devious still: he *wants* the rivals to *think* they are *overhearing* a command *intended* (of course) only for his own folk, and so on. Could a vervet possibly have that keen a sense of the situation? These dizzying heights of sophistication are strictly implied by the higher-order interpretation taken with its inevitable presupposition of rationality. Only a creature capable of appreciating these points could properly be said to have those beliefs and desires and intentions.

Another observation of the vervets brings out this role of the rationality assumption even more clearly. When I first learned that Seyfarth's methods involved hiding speakers in the brush and playing recorded alarm calls, I viewed the very success of the method as a seriously demoting datum, for if the monkeys really were Gricean in their sophistication, when playing their audience roles they should be perplexed, unmoved, somehow disrupted by disembodied calls issuing from no known utterer. If they were oblivious to this problem, they were no Griceans. Just as a genuine Communicator typically checks the Audience periodically for signs that it is getting the drift of the communication, a genuine Audience typically checks out the Communicator periodically for signs that the drift it is getting is the drift being delivered.

To my delight, however, I learned from Seyfarth that great care had been taken in the use of the speakers to prevent this sort of case from arising. Vervets can readily recognize the particular calls of their band – thus they recognize Sam's leopard call *as* Sam's, not Tom's. Wanting to give the recordings the best chance of "working," the experimenters took great care to play, say, Sam's call only when Sam was neither clearly in view and close-mouthed or otherwise occupied, nor "known" by the others to be far away. Only if Sam could be "supposed" by the audience to be actually present and uttering the call (though hidden from their view), only if the audience could *believe* that the noisemaker in the bush was Sam, would the experimenters play Sam's call. While this remarkable patience and caution are to be applauded as scrupulous method, one wonders whether they were truly necessary. If a "sloppier" scheduling of playbacks produced just as "good" results, this would in itself be a very important *demoting* datum. Such a test should be attempted; if the monkeys are baffled and unmoved by recorded calls except under the scrupulously maintained circumstances, the necessity of those circumstances would strongly support the claim that Tom, say, *does* believe that the noisemaker in the bush is Sam, that vervet monkeys are not only capable of believing such things, but *must* believe such things for the observed reaction to occur.

The rationality assumption thus provides a way of taking the various hypotheses seriously – seriously enough to test. We expect at the outset that there are bound to be grounds for the verdict that vervet monkeys are believers only in some attenuated way (compared to us human believers). The rationality assumption helps us look for, and measure, the signs of attenuation. We frame conditionals such as

14. If x believed that p, and *if x was rational,* then since "p" implies "q," x would (have to) believe that q.

This leads to the further attribution to x of belief that q,[3] which, coupled with some plausible attribution of desire, leads to a prediction of behavior, which can be tested by observation or experiment.[4]

Once one gets the knack of using the rationality assumption for leverage, it is easy to generate further telling behaviors to look for in the wild or to provoke in experiments. For instance, if anything as sophisticated as a third- or fourth-order analysis is correct, then it ought to be possible, by devious (and morally dubious!) use of the hidden speakers to create a "boy who cried wolf."[5] If a single vervet is picked out and "framed" as the utterer of false alarms, the others, being rational, should begin to lower their trust in him, which *ought* to manifest itself in a variety of ways. Can a "credibility gap" be created for a vervet monkey? Would the potentially nasty results (remember what happened in the fable) be justified by the interest such a positive result would have?

IV. How to use anecdotal evidence: The Sherlock Holmes method

One of the recognized Catch 22s of cognitive ethology is the vexing problem of anecdotal evidence. On the one hand, as a good scientist, the ethologist knows how misleading and, officially, unusable anecdotes are, and yet on the other hand they are often so telling! The trouble with the canons of scientific evidence here is that they virtually rule out the description of anything but the oft-repeated, oft-observed, stereotypic behavior of a species, and this is just the sort of behavior that reveals no particular intelligence at all – all this behavior can be more or less plausibly explained as the effects of some humdrum combination of "instinct" or tropism and conditioned response. It is the *novel* bits of behavior, the acts that couldn't plausibly be accounted for in terms of prior conditioning or training or habit, that speak eloquently of intelligence; but if their very novelty and unrepeatability make them anecdotal and hence inadmissible evidence, how can one proceed to develop the cognitive case for the intelligence of one's target species?

Just such a problem has bedeviled Premack and Woodruff (1978), for instance, in their attempts to demonstrate that chimps "have a theory of mind"; their scrupulous efforts to force their chimps into nonanecdotal, repeatable behavior that manifests the intelligence they believe them to have engenders the frustrating side effect of providing prolonged training histories for the behaviorists to point to in developing their rival, conditioning hypotheses as putative explanations of the observed behavior. [See the commentaries and replies in: "Cognition and Consciousness in Nonhuman Species" *BBS* 1(4) 1978; see also Premack: "The Codes of Man and Beasts" *BBS* 6(1) 1983.]

We can see the way out of this quandary if we pause to ask ourselves how we establish our *own* higher-order intentionality to the satisfaction of all but the most doctrinaire behaviorists. We can concede to the behaviorists that any single short stretch of human behavior can be given a relatively plausible and not obviously ad hoc demoting explanation, but as we pile anecdote upon anecdote, apparent novelty upon apparent novelty, we build up for each acquaintance such a biography of *apparent* cleverness that the claim that it is *all* just lucky coincidence – or the result of hitherto undetected "training" – becomes the more extravagant hypothesis. This accretion of unrepeatable detail can be abetted by using the intentional stance to provoke one-shot circumstances that will be particularly telling. The intentional stance is in effect an engine for generating or designing anecdotal circumstances – ruses, traps, and other intentionalistic litmus tests – and predicting their outcomes.

This tricky tactic has long been celebrated in literature. The idea is as old as Odysseus testing his swineherd's loyalty by concealing his identity from him and offering him temptations. Sherlock Holmes was a master of more intricate intentional experiments, so I shall call this the *Sherlock Holmes method.* Cherniak (1981) draws our attention to a nice case:

> In "A Scandal in Bohemia," Sherlock Holmes' opponent has hidden a very important photograph in a room, and Holmes wants to find out where it is. Holmes has Watson throw a smoke bomb into the room and yell "fire" when Holmes' opponent is in the next room, while Holmes watches. Then, as one would expect, the opponent runs into the room and takes the photograph from where it was hidden. Not everyone would have devised such an ingenious plan for manipulating an opponent's behaviour; but once the conditions are described, it seems very easy to predict the opponent's actions. (p. 161)

In this instance Holmes simultaneously learns the location of the photograph and confirms a rather elaborate intentional profile of his opponent, Irene Adler, who is revealed to *want* the photograph; to *believe* it to be located where she goes to get it; to *believe* that the person who yelled "fire" *believed* there was a fire (note that if she believed the yeller wanted to deceive her, she would take entirely different action); to *want* to retrieve the photograph without letting anyone *know* she was doing this, and so on.

A variation on this theme is an intentional tactic beloved of mystery writers: provoking the telltale move. All the suspects are gathered in the drawing room, and the detective knows (and he alone knows) that the guilty party (and only the guilty party) *believes* that an incriminating cuff link is under the gateleg table. Of course the culprit *wants* no one else to *believe* this, or to *discover* the cuff link, and *believes* that in due course it will be discovered unless he takes covert action. The detective arranges for a "power failure"; after a few seconds of darkness the lights are switched on and the guilty party is, of course, the chap on his hands and knees under the gateleg table. What else on earth could conceivably explain this novel and bizarre behavior in such a distinguished gentleman?[6]

Similar stratagems can be designed to test the various hypotheses about the beliefs and desires of vervet monkeys and other creatures. These stratagems have the virtue of provoking novel but interpretable behavior, of *generating anecdotes* under controlled (and hence scientifically admissible) conditions. Thus the Sherlock Holmes method offers a significant increase in investigative power over behaviorist methods. This comes out dramatically if we compare the actual and contemplated research on vervet monkey communication with the efforts of Quine's imagined behavioristic field linguist. According to Quine, a necessary preliminary to any real progress by the linguist is the tentative isolation and identification of native words (or speech acts) for "Yes" and "No," so that the linguist can enter into a tedious round of "query-and-assent" – putting native sentences to cooperative natives under varying conditions and checking for patterns in their yes and no responses (Quine 1960, chap. 2). Nothing just like Quine's game of query-and-assent can be played by ethologists studying animals, but a vestige of this minimalist research strategy is evident in the patient explorations of "stimulus substitution" for animal vocalizations – to the exclusion, typically, of more manipulative (if less intrusive) experiments (see note 1). So long as one is resolutely behavioristic, however, one must miss the evidential value of such behavior as the lone vervet quietly taking to the trees when a "leopard stimulus" is presented. But without a goodly amount of such telling behavior, no mountain of data on what Quine calls the "stimulus meaning" of utterances will reveal that they are communicative acts, rather than merely audible manifestations of peculiar sensitivities. Quine of course realizes this, and tacitly presupposes that his radical translator has already informally satisfied himself (no doubt by using the powerful, but everyday, Sherlock Holmes method) of the richly communicative nature of the natives' behavior.

Of course the power of the Sherlock Holmes method cuts both ways; failure to perform up to expectations is often a strongly demoting datum.[7] Woodruff and Premack (1979) have tried to show that chimpanzees in their lab can be full-fledged *deceivers*. Consider Sadie, one of four chimps used in this experiment. In Sadie's sight, food is placed in one of two closed boxes she cannot reach. Then either a "cooperative" or a "competitive" trainer enters, and Sadie has learned she must point to one of the boxes in hopes of getting the food. The competitive trainer, if he discovers the food, will take it all himself and leave. The cooperative trainer shares the food with Sadie. Just giving Sadie enough experience with the circumstances to assure her appreciation of these contingencies involves training sessions that give the behaviorist plenty of grist for the "mere reinforcement" mill. (In order to render the identities of the trainers sufficiently distinct, there was strict adherence to special costumes and rituals; the competitive trainer always wore sunglasses and a bandit's mask, for instance. Does the mask then become established as a simple "eliciting stimulus" for the tricky behavior?)

Still, setting behaviorists' redescriptions aside, will Sadie rise to the occasion and do the "right" thing? Will she try to deceive the competitive trainer (and only the competitive trainer) by *pointing to the wrong box?* Yes, but suspicions abound about the interpretation.[8] How could we strengthen it? Well if Sadie *really* intends to deceive the trainer, she must (being rational) start with the belief that the trainer does not already know where the food is. Suppose, then, we introduce all the chimps in an entirely different context to transparent plastic boxes; they *should* come to *know* that since they – and anyone else – can see through them, anyone can see, and hence come to *know*, what is in them. Then on a one-trial, novel behavioral test, we can introduce a plastic box and an opaque box one day, and place the food in the plastic box. The competitive trainer then enters, and lets Sadie see him looking right at the plastic box. If Sadie *still* points to the opaque box, she reveals, sadly, that she really doesn't have a grasp of the sophisticated ideas involved in deception. Of course this experiment is still imperfectly designed. For one thing, Sadie might point to the opaque box out of despair, seeing no better option. To improve the experiment, an option should be introduced that would appear better to her only if the first option was hopeless, as in this case. Moreover, shouldn't Sadie be puzzled by the competitive trainer's curious behavior? Shouldn't it bother her that the competitive trainer, on finding no food where she points, just sits in the corner and "sulks" instead of checking out the other box? Shouldn't she be puzzled to discover that her trick keeps working? She *should* wonder: Can the competitive trainer be that stupid? Further, better-designed experiments with Sadie – and other creatures – are called for.[9]

Not wanting to feed the galling stereotype of the philosopher as an armchair answerer of empirical questions, I will nevertheless succumb to the temptation to make a few predictions. It will turn out on further exploration that vervet monkeys (and chimps and dolphins, and all other higher nonhuman animals) exhibit mixed and confusing symptoms of higher-order intentionality. They will pass some higher-order tests and fail others; they will in some regards reveal themselves to be alert to third-order sophistications, while disappointing us with their failure to grasp some apparently even simpler second-order points. No crisp, "rigorous" set of intentional hypotheses of any order will be clearly confirmed. The reason I am willing to make this prediction is not that I think I have special insight into vervet monkeys or other species but just that I have noted, as anyone can, that much the same is true of us human beings. We are not ourselves unproblematic exemplars of third- or fourth- or fifth-order intentional systems. And we have the tremendous advantage of being voluble language users, beings that can be plunked down at a desk and given lengthy questionnaires to answer, and the like. Our very capacity to engage in linguistic interactions of this sort seriously distorts our profile as intentional systems, by producing illusions of much more definition in our operative systems of mental representation than we actually have (Dennett 1978a, chaps. 3, 16; Dennett 1981b). I expect the results of the effort at intentional interpretation of monkeys, like the results of intentional interpretations of small children, to be riddled with the sorts of gaps and foggy places that are inevitable in the interpretation of systems that are, after all, only imperfectly rational (see Dennett 1981a; 1981c; 1982).

Still, the results, for all their gaps and vagueness, will be valuable. How and why? The intentional stance profile or characterization of an animal – or for that matter, an inanimate system – can be viewed as what engineers would call a set of specs – specifications for a device with a certain overall information-processing *competence*. An intentional system profile says, roughly, *what information* must be receivable, usable, rememberable, transmittable by the system. It alludes to the ways in which things in the surrounding world must be represented – but only in terms of distinctions drawn or drawable, discriminations makable – and not at all in terms of the actual machinery for doing this work. (Cf. Johnston 1981 on "task descriptions.") These intentional specs, then, set a design task for the next sort of theorist, the representation-system designer.[10] This division of labor is already familiar in certain circles within artificial intelligence (AI);

what I have called the intentional stance is what Newell (1982) calls "the knowledge level." And, oddly enough, the very defects and gaps and surd places in the intentional profile of a less than ideally rational animal, far from creating problems for the system designer, point to the shortcuts and stopgaps Mother Nature has relied upon to design the biological system; they hence make the system designer's job easier.

Suppose, for example, that we adopt the intentional stance toward bees, and note with wonder that they seem to *know* that dead bees are a hygiene problem in a hive; when a bee dies its sisters *recognize* that it has died, and, *believing* that dead bees are a health hazard, and *wanting*, rationally enough, to avoid health hazards, they *decide* they must remove the dead bee immediately. Thereupon they do just that. Now if that fancy an intentional story were confirmed, the bee system designer would be faced with an enormously difficult job. Happily for the designer (if sadly for bee romantics), it turns out that a much lower order explanation suffices: dead bees secrete oleic acid; the smell of oleic acid turns on the "remove it" subroutine in the other bees; put a dab of oleic acid on a live, healthy bee, and it will be dragged, kicking and screaming, out of the hive (Gould & Gould 1982; Wilson, Durlach & Roth 1958).

Someone in artificial intelligence, learning that, might well say: "Ah how familiar! I know *just* how to design systems that behave like that. Shortcuts like that are my stock in trade." In fact there is an eerie resemblance between many of the discoveries of cognitive ethologists working with lower animals and the sorts of prowess mixed with stupidity one encounters in the typical products of AI. For instance, Roger Schank (1976) tells of a "bug" in TALESPIN, a story-writing program written by James Meehan in Schank's lab at Yale, which produced the following story: "Henry Ant was thirsty. He walked over to the river bank where his good friend Bill Bird was sitting. Henry slipped and fell in the river. Gravity drowned." Why did "gravity drown"? (!) Because the program used a usually reliable shortcut of treating gravity as an unmentioned *agent* that is always around pulling things down, and since gravity (unlike Henry in the tale) had no friends (!), there was no one to pull it to safety when it was in the river pulling Henry down.

Several years ago, in "Why Not the Whole Iguana?" (Dennett 1978c) I suggested that people in AI could make better progress by switching from the modeling of human microcompetences (playing chess, answering questions about baseball, writing nursery stories, etc.) to the whole competences of much simpler animals. At the time I suggested it might be wise for people in AI just to *invent* imaginary simple creatures and solve the whole-mind problem for them. I am now tempted to think that truth is apt to be both more fruitful, and, surprisingly, more tractable, than fiction. I suspect that if some of the bee and spider people were to join forces with some of the AI people, it would be a mutually enriching partnership.

V. A broader biological perspective on the intentional stance

It is time to take stock of this upbeat celebration of the intentional stance as a strategy in cognitive ethology before turning to some lurking suspicions and criticisms. I have claimed that the intentional stance is well suited to describe, in predictive, fruitful, and illuminating ways, the cognitive prowess of creatures in their environments, and that, moreover, it nicely permits a division of labor in cognitive science of just the right sort: field ethologists, given both their training and the sorts of evidence derivable by their methods, are in no position to frame – let alone test – positive hypotheses about actual representational machinery in the nervous systems of their species. That sort of hardware and software design is someone else's specialty.[11] The intentional stance, however, provides just the right interface between specialties: a "black box" characterization of behavioral and cognitive competences observable in the field, but couched in language that (ideally) heavily constrains the design of machinery to put in the black box.[12]

This apparently happy result is achieved, however, by the dubious decision to throw behaviorist scruples to the winds and commit acts of mentalistic description, complete with assumptions of rationality. Moreover, one who takes this step is apparently as unconcerned with details of physiological realization as any (shudder) dualist! Can this be legitimate? I think it will help to answer that question if we postpone it for a moment and look at adopting the intentional stance in the broader context of biology.

A phenomenon that will nicely illustrate the connection I wish to draw is "distraction display," the well-known behavior, found in many very widely separated species of ground-nesting birds, of feigning a broken wing to lure a predator that approaches the nest away from its helpless inhabitants (Simmons 1952; Skutch 1976). This seems to be *deception* on the bird's part, and of course it is commonly called just that. Its point is to *fool* the predator. Now if the behavior is *really* deceptive, if the bird is a real deceiver, then it must have a highly sophisticated representation of the situation. The rationale of such deception is quite elaborate, and adopting R. Dawkins's (1976) useful expository tactic of inventing "soliloquies," we can imagine the bird's soliloquy:

> I'm a low-nesting bird, whose chicks are not protectable against a predator who discovers them. This approaching predator can be *expected* soon to discover them unless I distract it; it could be distracted by its *desire* to catch and eat me, but only if it *thought* there was a *reasonable* chance of its actually catching me (it's no dummy); it would contract just that *belief* if I *gave it evidence that* I couldn't fly anymore; I could do that by feigning a broken wing, etc.

Talk about sophistication! It is unlikely in the extreme that any feathered "deceiver" is an intentional system of this intelligence. A more realistic soliloquy for any bird would probably be more along the lines of: "Here comes a predator; all of a sudden I feel this tremendous urge to do that silly broken-wing dance. I wonder why?" (Yes, I know, it would be wildly romantic to suppose such a bird would be up to such a metalevel wondering about its sudden urge.) Now it is an open and explorable empirical question just how sensitive a bird's cognitive control system is to the relevant variables in the environment; if birds engage in distraction display even when there is a manifestly better candidate for the predator's focus of attention (another, actually wounded bird or other likely prey, for instance), the behavior will be unmasked as very low order indeed (like the bees' response to oleic acid). If, on the other hand, birds – some birds anyway – exhibit considerable sophistication in their use of the stratagem (distinguishing different sorts of predators, or, perhaps, revealing appreciation of the fact that you can't fool the same predator with the same trick again and again), our higher-order interpretation of the behavior as genuinely deceptive will be promoted or even confirmed.

But suppose it turned out that the killjoy interpretation was closest to the truth; the bird has a dumb tropism of sorts and that's all. Would we thereupon discard the label "deception" for the behavior? Yes and no. We would no longer *credit the individual bird* with the rationale of deception, but that rationale won't just go away. It is too obvious that the raison d'être of this instinctual behavior is its deceptive power. That's why it evolved. If we want to know why this strange dance came to be provokable on just these occasions, its power to deceive predators will have to be distilled from all the myriad of other facts, known and unknown and unknowable, in the long ancestry of the species. But who *appreciated* this power, who *recognized* this rationale, if not the bird or its individual ancestors? Who else but Mother Nature herself? That is to say: nobody. Evolution by natural selection "chose" this design for this "reason."

Is it unwise to speak this way? I call this the problem of *free-floating rationales*. We start, sometimes, with the hypothesis that we can assign a certain rationale to (the "mind" of) some individual creature, and then we learn better; the creature is too stupid to harbor it. We do not necessarily discard the rationale; if it is no coincidence that the "smart" behavior occurred, we pass the rationale from the individual to the evolving genotype. This tactic is obvious if we think of other, nonbehavioral examples of deception. No one has ever supposed that individual moths and butterflies with eye spots on their wings figured out the bright idea of camouflage paint and acted on it. Yet the deceptive rationale is there all the same, and to say it is *there* is to say that there is a domain within which it is *predictive* and, hence, explanatory. (For a related discussion, see Bennett 1976, §§ 52, 53, 62.) We may fail to notice this just because of the obviousness of what we can predict: For example, in a community with bats but not birds for predators we don't expect moths with eye spots (for as any rational deceiver knows, visual sleight-of-hand is wasted on the blind and myopic).

The transmission of the rationale from the individual back to the genotype is of course an old trick. For a century now we have spoken, casually, of species "learning" how to do things, "trying out" various strategies; and of course the figurative practice has not been restricted to cognitive or behavioral traits. Giraffes stretched their necks, and ducks had the wisdom to grow webs between their toes. All just figurative ways of speaking, of course – at best merely dramatic expository shortcuts, one would think. But surprisingly, these figurative ways of speaking can sometimes be taken a lot more seriously than people had thought possible. The application of ideas from game theory and decision theory – for example, Maynard Smith's (1972; 1974) development of the idea of *evolutionarily stable strategies* – depended on taking seriously the fact that the long-term patterns in evolution figuratively described in intentional terms bore a sufficient resemblance to the patterns in short-term interactions between (rational) (human) agents to warrant the application of the same normative-descriptive calculi to them. The results have been impressive.

VI. The "Panglossian paradigm" defended

The strategy that unites intentional system theory with this sort of theoretical exploration in evolutionary theory is the deliberate adoption of *optimality models*. Both tactics are aspects of *adaptationism*, the "programme based on the faith in the power of natural selection as an optimizing agent" (Gould & Lewontin 1979). As Lewontin (1978b) observes, "optimality arguments have become extremely popular in the last fifteen years, and at present represent the dominant mode of thought."

Gould has joined his Harvard colleague Lewontin in his campaign against adaptationism, and they call the use of optimality models by evolutionists "the Panglossian paradigm," after Dr. Pangloss, Voltaire's biting caricature, in *Candide*, of the philosopher Leibniz, who claimed that this is the best of all possible worlds. Dr. Pangloss could rationalize any calamity or deformity – from the Lisbon earthquake to venereal disease – and show, no doubt, that it was all for the best. Nothing *in principle* could prove that this was not the best of all possible worlds.

The case leveled against adaptationist thinking by Gould and Lewontin has been widely misinterpreted, even by some of those who have espoused it, perhaps because of the curious mismatch between the rhetoric of Gould and Lewontin's attack and the mildness of their explicit conclusions and recommendations. They heap scorn on the supposed follies of the adaptationist mind set, which leads many to suppose that their conclusion is that adaptationist thinking should be shunned altogether. Their work was drawn to my attention, in fact, by critics of an earlier version of this paper who claimed that my position was a version of adaptationism, "which Gould and Lewontin have shown to be completely bankrupt." But when I turned to this supposed refutation of my fundamental assumptions, I found that the authors' closing summation finds a legitimate place in biology for adaptationist thinking. Theirs is a call for "pluralism," in fact, a plaint against what they see as an exclusive concentration on adaptationist thinking at the cost of ignoring other important avenues of biological thought. But still, the arguments that precede this mild and entirely reasonable conclusion seem ill-suited to support it, for they are clearly presented as if they were attacks on the fundamental integrity of adaptationist thinking, rather than support for the recommendation that we should all try in the future to be more careful and pluralistic adaptationists.

Moreover, when I looked closely at the arguments, I was struck by feeling of *déjà vu*. These arguments were not new, but rather a replay of B. F. Skinner's long-lived polemical campaign against "mentalism." Could it be, I wondered, that Gould and Lewontin have written the latest chapter of Postpositivist Harvard Conservatism? Could it be that they have picked up the torch that Skinner, in retirement, has relinquished? I doubt that Gould and Lewontin view the discovery of their intellectual kinship with Skinner with unalloyed equanimity,[13] and I do not at all mean to suggest that Skinner's work is the conscious inspiration for their own, but let us survey the extent of their agreement. [See also BBS special issue on the work of B. F. Skinner, forthcoming.]

One of the main troubles with *adaptationism*, Lewontin (1978b) tells us, is that it is too easy: "optimality arguments dispense with the tedious necessity of knowing anything concrete about the genetic basis of evolution," he remarks caustically; a healthy imagination is the only requirement for this sort of speculative "storytelling," and plausibility is often the sole criterion of such stories (Gould & Lewontin 1979, pp. 153–54).

One of the main troubles with *mentalism*, Skinner (1964) tells us, is "[mentalistic] way stations are so often simply invented. It is too easy." One can always dream up a plausible mentalistic "explanation" of any behavior, and if your first candidate doesn't work out, it can always be discarded and another story found. Or, as Gould and Lewontin (1979, p. 153) say about adaptationism, "Since the range of adaptive stories is as wide as our minds are fertile, new stories can always be postulated. And if a story is not immediately available, one can always plead temporary ignorance and trust that it will be forthcoming."[14]

Gould and Lewontin object that adaptationist claims are unfalsifiable; Skinner claims the same about mentalist interpretations. And both object further that these all too easy to concoct stories *divert attention* from the nitty-gritty hard details that science should look for: Gould and Lewontin complain that adaptationist thinking distracts the theorist from the search for evidence of nonadaptive evolution via genetic drift, "material compensation," and other varieties of "phyletic inertia" and architectural constraints; in Skinner's case mentalism distracts the psychologist from seeking evidence of histories of reinforcement. As Skinner (1971) complains, "The world of the mind steals the show" (p. 12).

Both campaigns use similar tactics. Skinner was fond of trotting out the worst abuses of "mentalism" for derision – such as psychoanalytic "explanations" (in terms of unconscious beliefs, desires, intentions, fears, etc.) of syndromes that turn out to have simple hormonal or mechanical causes. These are cases of gratuitous and incautious overextension of the realm of the intentional. Gould and Lewontin give as a bad example some sloppy jumping to conclusions by an adaptationist, Barash (1976), in his attempt to explain aggression in mountain bluebirds – the invention of an "anticuckoldry" tactic, complete with rationale, where a much simpler and more direct account was overlooked (Gould & Lewontin 1979, p. 154). They also "fault the adaptationist programme for its failure to distinguish current utility from reasons of origin," a criticism that is exactly parallel to the claim (which I have not found explicitly in Skinner, though it is common enough) that mentalistic interpretation often confuses post hoc rationalization with a subject's "real reasons" – which must be reformulated, of course, in terms of a prior history of reinforcement.

Finally, there is the backsliding, the unacknowledged concessions to the views under attack, common to both campaigns. Skinner notoriously availed himself of mentalistic idioms when it suited his explanatory purposes, but excused this practice as shorthand, or as easy words for the benefit of laymen – never acknowledging how much he would have to give up saying if he forswore mentalistic talk altogether. Gould and Lewontin are much subtler; they espouse "pluralism" after all, and both are very clear about the utility and probity – even the necessity – of *some* adaptationist explanations and formulations.[15] Anyone who reads them as calling for the extirpation, root and branch, of adaptationism seriously misreads them – though they decline to say how to tell a good bit of adaptationism from the bits they deplore. This is indeed a sharp disanalogy with Skinner, the implacable foe of "mentalism." But still, they seem to me not to acknowledge fully their own reliance on adaptationist thinking, or indeed its centrality in evolutionary theory.

This comes out very clearly in Gould's (deservedly) popular book of essays, *Ever since Darwin* (1977). In "Darwin's Untimely Burial" Gould deftly shows how to save Darwinian theory from that old bugbear about its reducing to a tautology, via a vacuous concept of fitness: "certain morphological, physiological, and behavioral traits should be superior *a priori* as designs for living in new environments. These traits confer fitness by an engineer's criterion of good design, not by the empirical fact of their survival and spread" (1977, p. 42).[16] So we can look at designs the way engineers do and rate them as better or worse, on a certain set of assumptions about conditions and needs or purposes. But that is adaptationism. Is it Panglossian? Does it commit Gould to the view that the designs selected will always yield the *best* of all possible worlds? The customary disclaimer in the literature is that Mother Nature is not an optimizer but a "satisficer" (Simon 1957), a settler for the near-at-hand *better*, the good enough, not a stickler for the *best*. And while this is always a point worth making, we should remind ourselves of the old Panglossian joke: the optimist says this is the best of all possible worlds; the pessimist sighs and agrees.

The joke reveals vividly the inevitable existence of a trade-off between constraints and optimality. What appears far from optimal on one set of constraints *may* be seen to be optimal on a larger set. The ungainly jury rig under which the dismasted sailboat limps back to port may look like a mediocre design for a sailboat until we reflect that given the conditions and available materials, what we are seeing may just be the best possible design. Of course it also may not be. Perhaps the sailors didn't know any better, or got rattled, and settled for making a distinctly inferior rig. But what if we allow for such sailor ignorance as a boundary condition? "Given their ignorance of the fine points of aerodynamics, this is probably the best solution *they* could have recognized." When do we – or must we – stop adding conditions? There is no principled limit that I can see, but I do not think this is a *vicious* regress, because it typically stabilizes and stops after a few moves, and for however long it continues, the discoveries it provokes are potentially illuminating.

It doesn't *sound* Panglossian to remind us, as Gould often does, that poor old Mother Nature makes do, opportunistically and short-sightedly exploiting whatever is at hand – until we add: she isn't perfect, but *she does the best she can*. Satisficing itself can often be shown to be the *optimal* strategy when "costs of searching" are added as a constraint (see Nozick 1981, p. 300 for a discussion). Gould and Lewontin are right to suspect that there is a tautology machine in the wings of the adaptationist theater, always ready to spin out a new set of constraints that will save the Panglossian vision – but they are, I think, committed to playing on the same stage, however more cautiously they check their lines.

Skinner is equally right when he insists that *in principle* mentalistic explanations are unfalsifiable; their logical structure *always* permits revision ad lib in order to preserve rationality. Thus if I predict that Joe will come to class today because he wants to get a good grade, and believes important material will be presented, and Joe fails to show up, there is nothing easier than to decide that he *must*, after all, have had some more pressing engagement, or not have known today's date, or simply have forgotten, or – a thousand other hypotheses are readily available. Of course maybe he was run over by a truck, in which case my alternative intentional interpretations are so much wheel spinning. The dangers pointed out by Skinner, and by Gould and Lewontin, are real. Adaptationists, like mentalists, do run the risk of building theoretical edifices out of almost nothing – and making fools of themselves when these card castles tumble, as they occasionally do. That is the risk one always runs whenever one takes the intentional stance, or the adaptationist stance, but it can be wise to take the risk since the payoff is often so high, and the task facing the more cautious and abstemious theorist is so extraordinarily difficult.

Adaptationism and mentalism (intentional system theory) are not *theories* in one traditional sense. They are stances or strategies that serve to organize data, explain interrelations, and generate questions to ask Nature. Were they theories in the "classical" mold, the objection that they are question begging or irrefutable would be fatal, but to make this objection is to misread their point. In an insightful article, Beatty (1980) cites the adaptationists Oster and Wilson (1978): " 'the prudent course is to regard optimality models as provisional guides to future empirical research and not as the key to deeper laws of nature' " (p. 312). Exactly the same can be said about the strategy of adopting the intentional stance in cognitive ethology.

The criticism of ever-threatening vacuity, raised against both adaptationism and mentalism, would be truly telling if in fact we always, or even very often, availed ourselves of the slack that is available in principle. If we were forever revising, post hoc, our intentional profiles of people when they failed to do what we expected, then the practice would be revealed for a sham – but then, if that were the case the practice would have died out long ago. Similarly, if adaptationists were always (or very often) forced to revise their lists of constraints post hoc to preserve their Panglossianism, adaptationism would be an unappealing strategy for science. But the fact about both tactics is that, in a nutshell, *they work*. Not always, but gratifyingly often. We are actually pretty good at picking the right constraints, the right belief and desire attributions. The bootstrapping evidence for the claim that we have in fact located all the important constraints relative to which an optimal design should be calculated is that we make that optimizing calculation, and it turns out to be predictive in the real world. Isn't this arguing in a circle? One claims to have located all the genuinely important constraints on the grounds that

1. the optimal design given those constraints is A
2. Mother Nature optimizes
3. A is the observed (that is, apparent) design.

Here one *assumes* Pangloss in order to infer the completion of one's list of constraints. What other argument could ever be used to convince ourselves that we had located and appreciated all the relevant considerations in the evolutionary ancestry of some feature? As R. Dawkins (1980, p. 358) says, an adaptationist theory such as Maynard Smith's evolutionarily stable strategy theory

> as a whole is not intended to be a testable hypothesis which may be true and may be false, empirical evidence to decide the matter. It is a tool which we may use to find out about the selection pressures bearing upon animal behavior. As Maynard Smith (1978) said of optimality theory generally: "we are *not* testing the general proposition that nature optimizes, but the specific hypotheses about constraints, optimization criteria, and heredity. Usually we test whether we have correctly identified the selective forces responsible."

The dangers of blindness in adaptationist thinking, pointed out so vividly by Gould and Lewontin, have their mirror image in any approach that shuns adaptationist curiosity. Dobzhansky (1956) says, in much the spirit of Gould and Lewontin, "The usefulness of a trait must be demonstrated, it cannot just be taken for granted." But, as Cain (1964) observes, "equally, its uselessness cannot be taken for granted, and indirect evidence on the likelihood of its being selected for and actually adaptive cannot be ignored. . . . Where investigations have been undertaken, trivial characters have proved to be of adaptive significance in their own right." Cain slyly compares Dobzhansky's attitude with Robert Hooke's curiosity about the antennae of insects in *Micrographia* (1665):

> What the use of these kind of horned and tufted bodies should be, I cannot well imagine, unless they serve for smelling or hearing, though how they are adapted for either, it seems very difficult to describe: they are in almost every several kind of Flies of so various a shape, though certainly they are some very essential part of the head, and have some very notable office assigned them by Nature, since in all Insects they are to be found in one or other form.

"Apparently," Cain concludes, "the right attitude to enigmatic but widely occurring organs was fully understood as long ago as the middle of the seventeenth century, at least in England" (1964, p. 50).

Finally, I would like to draw attention to an important point Gould makes about the *point* of biology, the ultimate question the evolutionist should persistently ask. This occurs in his approving account of the brilliant adaptationist analysis (Lloyd and Dybas 1966) of the curious fact that cicada reproductive cycles are prime-numbered-years long – 13 years, for instance, and 17 years: "As evolutionists, we seek answers to the question, why. Why, in particular, should such striking synchroneity evolve, and why should the period between episodes of sexual reproduction be so long?" (Gould 1977, p. 99). As his own account shows, one has not *yet* answered the why question posed when one has abstemiously set out the long (and in fact largely inaccessible) history of mutation, predation, reproduction, selection – with no adaptationist gloss. Without the adaptationist gloss, we won't *know why*.[17]

The contrast between the two sorts of answers, the scrupulously nonadaptationist, historic-architectural answer Gould and Lewontin *seem* to be championing, and the frankly Panglossian adaptationist answer one can also try to give, is vividly captured in one final analogy from the Skinnerian war against mentalism. I once found myself in a public debate with one of Skinner's most devout disciples, and at one point I responded to one of his more outrageously implausible Skinnerisms with the question, "Why do you say *that?*" His instant and laudibly devout reply was, "Because I have been reinforced for saying that in the past." My why-question asked for a justification, a rationale, not merely an account of historical provenance. It is just possible, of course, that any particular such why-question will have the answer: "no *reason* at all; I just happened to be caused to make that utterance," but the plausibility of such an answer drops to near zero as the complexity and *apparent* meaningfulness of the utterance rises. And when a supportable rationale for such an act is found, it is a mistake – an anachronistic misapplication of positivism – to insist that "the real reason" for the act *must* be stated in terms that make no allusion to this rationale. A purely causal explanation of the act, at the microphysical level, say, is *not in competition* with the rationale-giving explanation. This is commonly understood these days by postbehaviorist psychologists and philosophers, but the counterpart point is apparently not quite so well received yet among biologists, to judge from the following passage, in *Science*, reporting on the famous 1980 Chicago conference on macroevolution:

> Why do most land vertebrates have four legs? The seemingly obvious answer is that this arrangement is the optimal design. This response would ignore, however, the fact that the fish that were ancestral to terrestrial animals also have four limbs, or fins. Four limbs may be very suitable for locomotion on dry land, but *the real reason* [my emphasis] that terrestrial animals have this arrangement is because their evolutionary predecessors possessed the same pattern. (Lewin 1980, p. 886)

When biologists ask the evolutionists' why-question, they are, like mentalists, seeking the rationale that explains why some feature was selected. The more complex and apparently meaningful the feature, the less likely it is that there is *no* sustaining rationale; and while the historical and architectonic facts of the genealogy may in many cases loom as the most salient or important facts to uncover, the truth of such a nonadaptationist story does not *require* the falsehood of all adaptationist stories of the same feature. The *complete* answer to the evolutionists' question will almost always, in at least some minimal way, allude to *better* design.

Is this the best of all possible worlds? We shouldn't even try to answer such a question, but adopting Pangloss's assumption, and in particular the Panglossian assumption of rationality in our fellow cognizers, can be an immensely fruitful strategy in science, if only we can keep ourselves from turning it into a dogma.

ACKNOWLEDGMENTS

Among the many people who have advised me on earlier drafts, a few stand out: John Beatty, Colin Beer, Jonathan Bennett, Bo Dahlbom, Donald Griffin, Douglas Hofstadter, Harriet Kuliopoulos, Dan Lloyd, Ruth Garrett Millikan, David Policansky, David Premack, Carolyn Ristau, Sue Savage-Rumbaugh, Robert Seyfarth, Elliott Sober, Gabriel Stolzenberg, and Herbert Terrace. They are not responsible, of course, for the errors that remain – and one of the beauties of the *BBS* format is that they will have the opportunity to reveal that explicitly.

NOTES

1. We can probe the boundaries of the stimulus-equivalence class for this response by substituting for the "normal" leopard such different "stimuli" as dogs, hyenas, lions, stuffed leopards, caged leopards, leopards dyed green, firecrackers, shovels, motorcyclists. Whether these independent tests are tests of *anxiety specificity* or of the *meaning* of one-word sentences of Vervetese depends on whether our tests for the other components of our *n*th-order attribution, the nested intentional operators, come out positive.

2. See Quine (1960) pp. 48–49, on second-intention cases as "the bane of theoretical linguistics."

3. "I shall always treasure the visual memory of a very angry philosopher, trying to convince an audience that 'if you believe that A and you believe that if A then B then you *must* believe that B.' I don't really know whether he had the moral power to coerce anyone to believe that B, but a failure to comply does make it quite difficult to use the word 'belief,' and that is worth shouting about" (Kahneman 1982).

4. The unseen normality of the rationality assumption in any attribution or belief is revealed by noting that (14), which explicitly assumes rationality, is virtually synonymous with (plays the same role as) the conditional beginning: if *x really* believed that p, then since "p" implies "q" . . .

5. I owe this suggestion to Susan Carey, in conversation.

6. It is a particular gift of the playwright to devise circumstances in which behavior – verbal and otherwise – speaks loudly and clearly about the intentional profiles ("motivation," beliefs, misunderstandings, and so forth) of the characters, but sometimes these circumstances grow too convoluted for ready comprehension; a very slight shift in circumstance can make all the difference between utterly inscrutable behavior and lucid self-revelation. The notorious "get thee to a nunnery" speech of Hamlet to Ophelia is a classic case in point. Hamlet's lines are utterly bewildering until we hit upon the fact (obscured in Shakespeare's minimal stage directions) that while Hamlet is speaking to Ophelia, he *believes* not only that Claudius and Polonius are listening behind the arras, but that they *believe* he doesn't *suspect* that they are. What makes this scene particularly apt for our purposes is the fact that it portrays an intentional experiment: Claudius and Polonius, using Ophelia as decoy and prop, are attempting to provoke a particularly telling behavior from Hamlet in order thereby to discover just what his beliefs and intentions are; they are foiled by their failure to design the experiment well enough to exclude from Hamlet's intentional profile the belief that he is being observed, and the desire to create false beliefs in his observers. See, for example, Dover Wilson (1951). A similar difficulty can bedevil ethologists: "Brief observations of avocet and stilt behavior can be misleading. Underestimating the bird's sharp eyesight, early naturalists believed their presence was undetected and misinterpreted distraction behavior as courtship" (Sordahl 1981, p. 45).

7. I do not wish to be interpreted as *concluding* in this paper that vervet monkeys, or laboratory chimpanzees, or any nonhuman animals have *already been shown* to be higher-order intentional systems. Once the Sherlock Holmes method is applied with imagination and rigor, it may very well yield results that will disappoint the romantics. I am arguing in favor of a method of raising empirical questions, and explaining the method by showing what the answers *might be* (and why); I am not giving those answers in advance of the research.

8. It is all too easy to stop too soon in our intentional interpetation of a presumably "lower" creature. There was once a village idiot who, whenever he was offered a choice between a

dime and a nickel, unhesitatingly took the nickel – to the laughter and derision of the onlookers. One day someone asked him if he could be so stupid as to continue choosing the nickel after hearing all that laughter. Replied the idiot, "Do you think that if I ever took the dime they'd ever offer me another choice?"

The curiously unmotivated rituals that attended the training of the chimps as reported in Woodruff and Premack (1979) might well have baffled the chimps for similar reasons. Can a chimp wonder why these human beings don't just eat the food that is in their control? If so, such a wonder could overwhelm the chimps' opportunities to understand the circumstance in the sense the researchers were hoping. If not, then this very limit in their understanding of such agents and predicaments undercuts somewhat the attribution of such a sophisticated higher-order state as the desire to deceive.

9. This commentary on Premack's chimpanzees grew out of discussion at the Dahlem conference on animal intelligence with Sue Savage-Rumbaugh, whose chimps, Austin and Sherman, themselves exhibit apparently communicative behavior (Savage-Rumbaugh, Rumbaugh & Boysen 1978) that cries out for analysis and experimentation via the Sherlock Holmes method.

10. In the terms I develop in "Three Kinds of Intentional Psychology" (Dennett 1981b), intentional system theory specifies a semantic engine which must then be realized – mimicked, in approximation – by a syntactic engine designed by the subpersonal cognitive psychologist.

11. I should acknowledge, though, that in the case of insects and spiders and other *relatively* simple creatures, there are some biologists who have managed to bridge this gap brilliantly. The gap is much narrower in nonmammalian creatures, of course.

12. In "How to Study Consciousness Empirically: Or, Nothing Comes to Mind" (Dennett 1982), I describe in more detail how purely "semantic" descriptions constrain hypotheses about "syntactic" mechanisms in cognitive psychology.

13. For all their manifest differences, Lewontin and Skinner do share a deep distrust of cognitive theorizing. Lewontin (1981) closes his laudatory review of Gould's *The Mismeasure of Man* (1981) in the *New York Review of Books* with a flat dismissal of cognitive science, a verdict as sweeping and undiscriminating as any of Skinner's obiter dicta: "It is not easy, given the analytic mode of science, to replace the clockwork mind with something less silly. Updating the metaphor by changing clocks into computers has got us nowhere. The wholesale rejection of analysis in favor of obscurantist holism has been worse. Imprisoned by our Cartesianism, we do not know how to think about thinking" (p. 16).

14. This objection is familiar to E. O. Wilson (1975), who notes: "Paradoxically, the greatest snare in sociobiological reasoning is the ease with which it is conducted. Whereas the physical sciences deal with precise results that are usually difficult to explain, sociobiology has imprecise results that can be too easily explained by many different schemes" (p. 20). See also the discussion of this in Rosenberg (1980).

15. Lewontin, for instance, cites his own early adaptationist work, "Evolution and the Theory of Games" (Lewontin 1961), in his recent critique of sociobiology, "Sociobiology as an Adaptationist Program" (Lewontin 1979). And in his *Scientific American* article (Lewontin 1978a), "Adaptation," he concludes: "to abandon the notion of adaptation entirely, to simply observe historical change and describe its mechanisms wholly in terms of the different reproductive success of different types, with no functional explanation, would be to throw out the baby with the bathwater" (p. 230).

16. For a more rigorous discussion of how to define fitness so as to evade tautology, see Rosenberg (1980, pp. 164–75).

17. Boden (1981) advances the claims for "the cognitive attitude" (in essence, what I have called the intentional stance) in a different biological locale: the microstructure of genetics, enzyme "recognition" sites, embryology, and morphogenesis. As she says, the cognitive attitude "can encourage biologists to ask empirically fruitful questions, questions that a purely physico-chemical approach might tend to leave unasked" (p. 89).

Open Peer Commentary

Commentaries submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Rationality: Putting the issue to the scientific community

John Beatty
Department of Philosophy, Arizona State University, Tempe, Ariz. 85287

Vervet rationality aside for the moment, what about scientific rationality? One of the main questions Dennett raises in this regard concerns the methodological soundness of persisting in the search for certain kinds of scientific explanations rather than others – in particular, rationalist versus behaviorist explanations in ethology, and selectionist versus alternative evolutionary explanations. As Dennett explains, rationalist and selectionist explanations have in common that they are optimal-solution explanations. What Dennett succeeds pretty well in doing, in defense of such pursuits, is to relax a methodological stricture – a kind of pluralistic demand – that enjoins rationalists, selectionists, and other optimizers to pursue alternative explanations sometimes prior to their optimal-solution pursuits, or at least concurrently with those pursuits, and certainly after proposed optimal-solution accounts have been falsified (there are many ways of interpreting the pluralist appeal, as Dennett discusses in the context of Gould and Lewontin's version). Dennett argues that there are conditions under which it is not unreasonable to pursue optimal-solution explanations of ethological and biological phenomena, practically to the exclusion of alternative types of explanation, even after proposed optimal-solution accounts of those phenomena have been refuted.

I said that Dennett "succeeds pretty well," when what I meant is that he succeeds well at a certain level. This discussion can, and, I think, should, be carried on at another level as well – a level recently brought to light by Sarkar (1982). As Sarkar points out, there is a distinction between the principles that govern the rationality of individual scientists and the principles that govern the rationality of communities of scientists. A principle may apply at one level and not the other. Of particular relevance to the issue at hand, Sarkar suggests that pluralistic methodological appeals may be more appropriately aimed at communities than at individual scientists.

Dennett seems to me to be considering methodological pluralism only as a maxim to which individual scientists should or should not adhere. But in so doing, he overlooks a more appealing version of methodological pluralism. I read (construe?) pluralistic methodological critiques of hard-line optimal-solution approaches as appeals to the community. When Gould and Lewontin complain about the prevalence of selectionist approaches in evolutionary biology, for instance, I consider the target of their complaint to be for the most part the *community* of evolutionary biologists. Individual evolutionary biologists are not so much their targets, but are instead the only kinds of instances to which they have recourse, in order to point out the selectionist excesses of the community. Their point is, in other words, not that it is unreasonable for an individual scientist to be a stalwart selectionist (indeed, they respect the diligence of selectionists like Paul Sheppard – see, e.g., Lewontin's 1972 references to Sheppard), but that it is unreasonable for the community of evolutionary biologists to put all its eggs into the selectionist basket. The community of evolutionary biologists should support the pursuit of a variety of evolutionary accounts – not just, and perhaps not even primarily, selectionist accounts. So it's not so much that each evolutionary biologist should pursue a plurality of approaches, as that the community should consist of a variety of different kinds of evolutionary biologists. The same goes, I would presume, for the community of ethologists. It may not be unreasonable for any individual ethologist to be either a hard-line behaviorist or a hard-line rationalist. But it would be unreasonable, at this time, for the community of ethologists to support either approach to the exclusion of the other.

Conceding, in other words, Dennett's defense of the reasonableness of being a stalwart rationalist in the face of behaviorist objections, and of being a stalwart selectionist in the face of alternative evolutionary objections, the question still remains as to how the communities of ethologists and evolutionists should divide up their support among the variety of approaches open to their members. How much of those communities' efforts should be devoted to optimal-solution approaches?

Cognitive ethology: Theory or poetry?

Jonathan Bennett

Department of Philosophy, Syracuse University, Syracuse, N.Y. 13210

Dennett is perhaps the most interesting, fertile, and challenging philosopher of mind on the contemporary scene, and I count myself among his grateful admirers. But this present paper of his, enjoyable as it is to read, and acceptable as its conclusions are, is likely to do more harm than good. Some will object that the intentional stance is a dead end; but I think, as Dennett does, that it is premature to turn our backs on explanations of animal behavior in terms of desires and beliefs, and I am in favor of continuing with this endeavor; but only if it gets some structure, only if it is guided by some firm underlying theory. That is what the ethologists might get from philosophy, but Dennett has invited them to turn their attention toward philosophy only to give them a mildly upgraded version of the unstructured, opportunistic, rambling kind of thing they are doing already. He encourages them to go on believing that the conceptual foundations of cognitive ethology are rather easy to lay – a few broad strokes of the brush, or slaps of the trowel, and there you are. Really, it is much harder and more laborious than that. I shall sketch the sort of thing that is needed, and point out some things in Dennett's paper that suffer from the lack of any proper foundations.

I take it as uncontroversial that the intentional stance – considered as a program for theorizing about behavior – must be centered on the idea that beliefs are functions from desires to behavior, and that desires are functions from beliefs to behavior. Down in the foundations, then, we need some theory about what behavior must be like to be reasonably interpreted as manifesting beliefs and desires; and these concepts must presumably tail off somehow, being strongly applicable to men and apes, less strongly to monkeys, and so on down to animals that do not have beliefs and desires but can be described in terms of weaker analogues of those notions. What will this "tailing off" look like? In my attempt to answer this (Bennett 1976) I have taken it that a theory of belief and desire will be nested within a broader theory of *goals*, or of a *teleological explanation of behavior*. The basis for the latter is to be found, I think, in Taylor (1964), which highlights the idea of what I call an "instrumental property" of an organism, that is, a property of the form: "x is so situated and constructed that if it soon does A it will become G shortly thereafter." Let us put this by saying, for short, that the animal is A/G.

Teleological explanations come into play only if we have a system (e.g. an animal) regarding which there is a reliable generalization of the form: "For any *a*, whenever it is *a*/G it proceeds to do *a* if that is within its physical competence." If the animal has eating as its G, its goal, it will dependably kill when it is killing/eating, climb when it is climbing/eating, and so on. I am suppressing many complications, but one must be faced openly. No actual animal will, for any G, do whatever *will* bring it G. You give me an animal and a value of G for which this is supposed to be true, and I will rig a situation in which the animal will get G if and only if it lies down and then stands up, three times in quick succession (e.g. I will *decide* to give it G if and only if it behaves in that way). But it won't act like that unless I somehow inform it of the relevant fact about its situation. So a theory of teleology that is to have any chance of fitting actual animals must rest on generalizations not of the form "If it *is a*/G it will do *a*" but rather "If it *has the information that it is a*/G it will do *a*."

In my book I coined the term "registration," speaking of the animals's being *a*/G as a fact that may be "registered" upon it; and then I argued that belief is a species of registration, the differential being a matter of degree which I tried to describe. I probably didn't get it right, but that is of no great moment. What marks off the genus "goal" from the species "desire" or "intention," and the genus "registration" from the species "belief," is far less important than is the structure of the genus. That is, what matters is to have a good theory of teleologically explicable behavior, with the foundations of a theory of cognition embedded in it. And I offer my attempt at this in Bennett (1976) not as a source of the right answers, perhaps, but as a fair indication of what some of the principal questions are. I contend that something of that general nature – and not less complex than that – is needed as a foundation for the intentional stance, if the latter is to be worth anything as theory, rather than merely expressing a liking for one way of talking, a kind of dim poetry.

The most important thing in any foundational theory will be its answer to the question, What makes it all right to explain an event teleologically, bringing it under a generalization of the broad form of "If x registers that it is *a*/G it does *a*, for any *a* within its physical competence"? If the event could be explained in that way and no other, that would justify using the teleogical explanation. But what if every event can be explained mechanistically, that is, in terms of its subject's intrinsic properties, with no mention of any property of the form A/G? I answer that it is all right to bring x under a teleological generalization if the latter captures a class of events that is not covered by any *one* generalization of a mechanistic sort. Where there is a contest between one teleological and one mechanistic generalization (or even, perhaps, two or three of the latter), mechanism wins because it is more basic, uses concepts of wider applicability, and so on (see Taylor 1964, p. 29). But if a teleological generalization does work for us – giving us classifications, comparisons, contrasts, patterns of prediction that mechanism does not easily provide – then that justifies us in employing it. This, I submit, is the *Grundgesetz* of the whole theory of teleological explanation and thus of the intentional stance.

It bears heavily on one of Dennett's themes. He rightly says that any attribution of beliefs and the like to an animal must be able to stand its ground against lower-level "killjoy" rivals; and he gives some nice examples of attributions withdrawn – or behavior "demoted" – in the light of further evidence. This can happen not only when a high-level intentional attribution is challenged by a lower-level one ("Does he want me to think he is hungry, or only to give him food?") but also when a lowest-level intentional attribution is challenged by a rival that does not involve intentionality at all.

There is a problem about the latter kind of issue, which Dennett describes but does not explain. Suppose we are inclined to think that a certain animal has as a goal escaping from leopards. That is, whenever its being *a*/escapes-from-leopard is registered on it, it does *a* (subject to complications and qualifications which I shall continue to omit). What would a challenge from below, a killjoy rival, look like in such a case? It would consist in the discovery that the class of events that we had brought under a teleological generalization could also be brought under a nonteleological one. For purposes of this particular point I shall simplify the teleological form even further, and take it to be: "If the animal (registers that it) is in a leopard-threatening situation it does a leopard-avoiding thing." We might opt for that generalization – or for the teleological one of which it is a simplified caricature – because we could find no principle of unity for that class of events except the one provided by "leopard-betokening" in the input and "leopard-avoiding" in the output. But now suppose we discovered that there is a kind of stimulus S and a kind of behavior R such that (i) S is definable without help from any concept like that of "being evidence for" or "registering" (e.g. S is a kind of smell, definable in purely chemical terms), and (ii) R is definable without help from any concept like that of "tending to" or "being apt for" (e.g. R is a motor kind of movement, definable in terms of how certain muscles are used), and (iii) the class of supposedly leopard-avoiding situations also falls under the generalization that whenever the animal receives an S stimulus it emits an R response. In that case, the generalization "Whenever it is in (what it registers as being) a leopard-threatening situation it does a leopard-avoiding thing" should be relinquished: The intentional stance has no honest work to do here, because all its work is equally done by something that is preferable to it because lower level. (Whether the S-R pattern is hard-wired or a result of learning is quite irrelevant, so far as I can see.)

Now, Dennett sees intentional explanations of behavior as threatened by stimulus-response rivals, but he does not say why, except to remark that "the acts that couldn't plausibly be accounted for in terms of prior conditioning or training or habit [are the ones] that speak eloquently of intelligence" and thus of intentionality. If Dennett wants to be really useful to cognitive ethologists and psychologists – giving them what they need rather than what they want – he ought not to be talking in this way about what "speaks eloquently" of what, nor should he rely on the term "training," trusting his intended audience to understand how the kind of training that does not require intentionality differs from the kind of learning that does. Rather, he should be helping them to understand what conceptual *structures* are involved here. That would require him to have much more theory than he has. He would have to descend from the level of sweeping remarks about stances and levels, and talk in detail about how the levels relate to one another.

This lack of theoretical structure goes very deep in Dennett's paper – right down to the level of the question of what intentionality *is*. Apart from giving its nominal essence by saying that it is the home ground of intentions, beliefs, and the like, Dennett mentions only one thing that can "mark" the sphere of the intentional, namely that it involves referential opacity. But the converse doesn't always hold: Some opaque contexts are not intentional; and in any case, how is our grasp of intentionality supposed to be *helped* by this mention of opacity? It has nothing to offer to the foundering ethologist or psychologist, and Dennett makes no use of it in the subsequent discussion. He did need to say something of a technical nature about intentionality, but not *that*. What was needed was rather an account of intentionality as the locus of one kind of function from sensory inputs to behavioral outputs of animals: A description of what those functions are, of how they actually work, would have meshed with things that ethologists and psychologists do, helping them to get somewhere with their problems; whereas what Dennett says about opacity does not turn any of the wheels that badly need turning at present.

The absence from Dennett's paper of any theoretical underlay also makes itself felt in his treatment of what he sees as a problem confronting anyone who wants to base intentionalist conclusions on ethological data. The problem, according to Dennett, is that the best evidence for intentionality comes from what an animal does in unusual circumstances; such evidence will take the form of relatively isolated anecdotes; and trained observers are taught to be wary of anecdotes, and to concentrate on getting hard data, that is, oft-repeated patterns of behavior. So there is a danger that accepted canons of good scientific conduct will act as a sieve, keeping the best evidence for intentionality from getting through onto the pages of the observer's log book. For this difficulty, he offers two solutions: (i) We can "pile anecdote upon anecdote, apparent novelty upon apparent novelty," until it becomes incredible that there is not a real underlying intentional pattern. (ii) We can devise experiments, set traps, and so on, trying to provoke "novel but interpretable behavior," thus "*generating anecdotes* under controlled (and hence scientifically admissible) conditions."

I object that Dennett has not explained why the problem exists, because he has not said why nonanecdotal evidence cannot support attributions of belief and desire, except for remarking that it does not "speak eloquently" of intentional states and may be explainable in terms of "conditioning or training or habit." I also object that he does not explain why his proposed solutions *are* solutions, or, for that matter, how they are to be executed. He does not say what kind of "pile" we should heap up in solution (i), and in (ii) he leaves it unclear how the poison of anecdote is supposed to be neutralized by the antidote of control.

In fact, his problem arises only if observers are looking for behavior that can be brought under generalizations relating *sensory* kinds of input to *motor* kinds of output, for example, saying that when the animal encounters a certain kind of smell it moves certain muscles thus and so, rather than generalizations relating *evidential* kinds of input to *consequential* kinds of output, for example, saying that when the animal encounters signs of the proximity of a leopard it does something that is apt to get it out of the leopard's vicinity. Suppose we have an animal that whenever it encounters an S smell, makes R movements; and suppose that usually an S smell is evidence of leopards and R movements do provide escapes from leopards. Now, we are wondering whether this behavior, conforming as it does to an S-R pattern, should be explained intentionally, that is, brought under the generalization that when the animal is (or perceives itself as) leopard threatened it leopard avoids. To find the answer, we must vary the conditions, bringing it about that the animal sometimes gets evidence of leopards other than S smells, and sometimes needs something other than an R movement to avoid a leopard; and we must observe how it behaves in these situations, either on a first encounter or after a number of trials from which the animal can learn things about evidence for leopards and means of escape from them. If the "leopard" generalization holds good in cases in which the S-R generalization fails, or in cases in which it is inapplicable, that helps to justify our using the "leopard" generalization, which is tantamount – given the simplification with which I am now working – to bringing the intentional stance to bear on the behavior in question. Despite what Dennett says, this is not a move from regularities to anecdotes; rather, it is a move from regularities of one kind to regularities of another. If the work is done right, there may indeed be "control," but that is not what makes the procedure "scientifically admissible." There is no reason in principle why we should not make the enlarged set of observations with our hands behind our backs, not contriving anything but just looking in the right direction. The procedure is scientifically admissible just because it consists in objectively attending to data in the light of a decent hypothesis; and it bears on intentionality because of what the hypothesis is. I think, as Dennett evidently does, that the ethological and psychological literature contains little convincing evidence of nonhuman intentionality. But that is not because intentionality is inimical to regularity and thus to normal scientific method; rather, it is because the people doing the work don't know what regularities to look for, having no theory of intentionality. I am afraid that Dennett's paper will encourage them to go on being content to have none.

Theoretical foundations are needed not only along the borderline between intentional and nonintentional, but also in adjudicating between a given intentional hypothesis and some lower-level intentional rival to it. Consider, for example, the contrast between "Tom wants Sam to believe that there is a leopard" and "Tom wants Sam to run into the trees." Dennett rightly implies that behavioral evidence can discriminate between these, but his only suggestion about how it can do so is wrong or seriously incomplete. He handles "Tom wants Sam to run into the trees" in terms of Tom's using a "trick" to "induce a certain response in Sam," and compares this with getting someone to jump by shouting "Boo!" at him. The impression is given that a first-order intention must be an intention to trigger an automatic response; but that is just wrong, for we have a first-order intention whenever an animal intends to bring it about that P, where P does not involve any intentional concepts. Thus, Tom may intend to get Sam to run into the trees, and the mechanism that actually operates in Sam may involve an inference from "Tom wants me to run to the trees, and usually it pays to do what Tom wants me to do" to the conclusion "It will be worthwhile to run to the trees." Tom's intentionality is not prevented from being first order by the fact that what happens in Sam – as distinct from what Tom intends or wants to happen in Sam – is itself intentional.

How, then, *can* behavior mark the difference between "wants Sam to believe there is a leopard" and "wants Sam to run into the trees"? Well, I think that it cannot mark the difference unless there are circumstances in which Tom thinks there is a leopard nearby and in which that fact makes it appropriate (relative to Tom's value system) for Sam to do something other than running to the trees. If there is a kind of behavior that Tom engages in whenever he thinks there is a leopard nearby, and if in each instance he behaves with the intention of getting Sam to do A, or do B, or do C, through a long list of kinds of behavior that have nothing in common except their appropriateness to there being a leopard nearby, *then*, and only then, are we entitled to say that what Tom wants is something describable with the aid of "There is a leopard in the vicinity." (I am here applying some thoughts I first developed in Bennett 1964, pp. 19–21.) It may, however, only be "Tom wants Sam to do something appropriate to the fact that there is a leopard in the vicinity." To be entitled to say that Tom wants Sam to *believe* that there is a leopard, we shall need further evidence; and it won't be easy to find. I suspect, indeed, that if we are ever to be entitled to interpret nonhuman animals in terms of anything higher than first-order intentionality, that will have to be because for nonhuman animals we adopt specially weakened intentional concepts. But perhaps not. In any case, whatever concepts are being used, they had better be understood; they had better exist as theoretical items, not as mere predilections for using words in certain ways. Otherwise the entire project will continue to wander in the wilderness.

I wonder what Dennett's picture is of the project as it has been pursued up to now. In a footnote he refers to two of the Yerkes chimpanzees, saying that their "apparently communicative behavior . . . cries out for analysis and experimentation via the Sherlock Holmes method," that is, through the accumulation of controlled and contrived anecdotes. He does not mention the fact that the Yerkes psychologists think that they *have* analyzed the communicative behavior of their chimpanzees (Savage-Rumbaugh, Rumbaugh & Boysen 1978). He must think it is possible to do better than they have, and I agree with that. But what kind of improvement in the analysis does Dennett have in mind? He gives the impression of thinking that adjudicating between rival interpretations is always to be handled in terms of informal, intuitive intelligence as brought to bear on the particular case, and that it shouldn't be very hard to get agreement by such means. If that is his view, then presumably he will think that what philosophy has to offer is just some rough guidance on how to be "careful" in thinking about cases, some help in getting "the knack." I submit that that is far too undemanding a picture of what is needed in adjudicating between rivals. And even if it were not, a proper underlying theory would still be needed to help students of animal behavior to know how to construct rivals to a given hypothesis and how to look for positive evidence that there aren't any rivals. Both sorts of help are desperately needed, judging by the literature to date.

Dennett's handling of the intentional stance – typified by his willingness to describe data in terms of what "speaks eloquently" and what "delights" or "depresses," rather than of what does or does not satisfy explicitly stated criteria – is puzzling. For he declares an interest in developing a "suitably rigorous abstract language in which to describe cognitive competences," and says: "We are interested in asking what gains in perspicuity, in predictive power, in generalization, might accrue if we adopt a higher-level hypothesis that takes a risky step into intentional characterization." Despite a puzzling later remark about the stance as not a theory "in one traditional sense," he clearly does regard it as enough of a theory to make my criticisms prima facie relevant. I can only suppose that his silence about all the theory's details arises from his thinking that the details are rather obvious and easy. Well, they are not; and much more work must be done on them if the intentional stance is to get anywhere. In implying the contrary, Dennett has misestimated the confusion and conceptual shallowness that reign throughout the relevant literature to date. And he has also misestimated his own needs in this very paper, as I have tried to show in pointing to some (not all) of the things in the paper that would have gone better if some explicit theory had been at work.

Dennett's instrumentalism: A frog at the bottom of the mug

Patricia Smith Churchland
Institute for Andvanced Study, Princeton, N.J. 08540

Apparently it makes good sense to think of animals as having sentential attitudes such as beliefs and desires. But behind the assurances I sense troubles. Does Dennett think that vervets have beliefs and desires in the way humans do? That is, if I have the belief that there is a leopard nearby, is that, in Dennett's view, the same sort of internal state the vervet is in when he has the belief that a leopard is nearby? Assuming "there is a leopard nearby" is a mental representation that both the vervet and I have, how should mental representations be characterized? In seeking answers to these questions, we find that the solid, practical, get-on-with-it character of common sense begins to look decidedly fragile, problematic, and sticky. For one thing, it turns out that within the cognitive framework the best theory going of how to characterize mental representations construes them as *sentences*, and sees someone's having a mental state, such as a belief, as the person standing in relation to a sentence in his inner language. On this theory, cognitive processes are understood to be manipulations of *linguistic structures* (see Fodor 1975; 1980). While there are certain advantages to the theory, the idea that, in general, representations are to be modeled on linguistic structures is dreadfully hard to stomach. For example, it entails the postulation of an innate and inner language, Mentalese, in virtue of which the individual believes, reasons, acquires his spoken language, and so on. Notice that if animals are ascribed beliefs and desires, then on this conception of representations, they too have an innate and inner language, and *their reasoning consists in an inner manipulation of linguistic symbols*. Accordingly, if the vervet and I both believe there is a leopard nearby, we both stand in a particular relation to the same inner linguistic structure, one analogous to the English sentence "there is a leopard nearby."

I find this view of representations excessively narrow and impossible to accept, for humans or for other animals (see P. S. Churchland 1980a; 1980b; see also Stich 1982). While there are a host of serious objections, the point of emphasis here is that human language has the earmarks of a device for *communication*, and as such, it is probably a latecomer in the representational business of organisms. Evolutionary and neurobiological considerations cry out against the likelihood that all representations in the brains are to be modeled on linguistic representations; rather, the deeper understanding of linguistic representation may come from a theory of more primitive kinds of representations. The implication that vervet or rattlesnake brains manipulate sentences, noun phrases, verb phrases, and the rest, is grimly far-fetched, though *representations* they surely do enjoy (see P. S. Churchland & P. M. Churchland 1983). In short, the best cognitively based theory of representations looks profoundly troubled.

How does Dennett cope with this problem? First, he agrees that the sentences-in-the-head theory of beliefs is a flop; indeed, he has argued this most convincingly himself (Dennett 1978a). Second, he nonetheless sees no problem in ascribing beliefs to humans and to vervets. But then what structures does Dennett think are in the head such that the vervet and I both believe there is a leopard nearby? His answer: none. Beliefs, it appears, are not (for the most part?) in the head; they are "virtual" (Dennett 1981b, pp. 48–49), and the actual, causally relevant states in the production of behavior are not beliefs and desires at all. This should be deeply troubling, for if beliefs and desires are not really in the brain, then it becomes questionable whether any explanatory work is done by attributing beliefs and desires to organisms. It appears that Dennett, faced with the catastrophe devolving from the sentences-in-the-head theory of cognitive representations, has engaged in special pleading for belief-desire psychology: he claims that belief-desire theory cannot be falsified, and that it is merely a way of organizing data, as opposed, presumably, to postulating causally active inner states of the organism.

The crux of the matter is that Dennett gives belief-desire psychology an *instrumentalistic* interpretation, and beliefs and desires are therefore no longer construed as real states of the brain, but as virtual states. He argues that belief-desire "theory" is not a theory in the "classical" mode; but then, alas, it certainly does not permit *explanations* in any recognizable mode either, for, as an instrument, it does not explain at all. As a convenient fiction, it may help to predict, but as a *fiction*, it cannot pretend to any explanatory function. One can decide to be an instrumentalist about any theory: vitalism, alchemy, Aristotelian biology, and so on. Just claim that the theory does not aspire to truth, but is a convenient fiction, and so cannot be falsified either. But there is no acceptable way to be an instrumentalist while playing the explanatory game. Part of the price you pay for saving a theory from revision by giving it instrumentalistic shelter and "protected status" is that the theory is thereby put beyond the bounds of testability, falsifiability, reducibility, coherence with related theories, and general scientific status. That appears not to bother Dennett at all, but once pointed out, it may reduce any ethologist's enthusiasm for Dennett's approach. Of course, there is not yet a theory available with which to replace belief-desire psychology, so we muck on nonetheless. Nor do I object to using a flawed theory to make do, especially if it can be the means of bootstrapping up to an entirely new theory, so long as the absence of a replacing theory is not used to sanctify the doddering theory. But by propping up belief-desire psychology with an instrumentalist stick, one fosters the illusion that it can stand on its own feet like any scientific theory, with the result that the impetus to improve upon it, revise it, and look for alternative theories, is diminished.

Having thus scowled and scoffed, I should nevertheless like to applaud Dennett's underlying conviction that human and nonhuman behavior alike ought to be encompassed by a single theory. Otherwise put, *if* belief-desire psychology is an adequate theory for human behavior, then it should be adequate to lots of nonhuman behavior. My point is that belief-desire psychology is fraught with troubles, and ethologists should be warned that if they buy into it, they buy into its troubles as well (P. M. Churchland 1981).

Science as an intentional system

Arthur C. Danto

Department of Philosophy, Columbia University, New York, N.Y. 10027

Someone who seeks to deny in his heart the existence of intentional systems falls, like Saint Anselm's fool, into self-stultification, since rational denial exemplifies the working of an intentional system. Scientists themselves, in the activities that define them as such, so conspicuously exemplify the working of intentional systems that if explanatory reference to these begs a question, the question of science itself is begged. These dialectical self-entrapments constitute a paraontological proof for the existence of intentional systems, the question being only what else in the universe save fools and scientists exemplifies them. I wish in this commentary to defend Daniel Dennett's thesis by demonstrating that the structure of the controversy in which he engages with those who deny intentional systems presupposes precisely what is at issue, so there is no further issue to discuss. There are only cases. My sole quarrel with Dennett is with his insistence that "intentional system theory" is but a *stance*, hence not a theory "in one traditional sense." The only sense in which it is not a theory is that in which "theory" contrasts with "fact." For if it were not fact it could not be discussed.

1. The predictable grudgingness of ideologized behaviorism toward countenancing intentionalized descriptions of mere monkeys is perhaps minor by comparison with rejecting such descriptions of its own practitioners. If chairholders in psychology departments and heads of laboratories may have their conduct fully represented in the ascetic idiom of operant conditioning, why suddenly become profligate in portraying the phobic conduct of simians? There is an admirable consistency in this resolution to bed down with the troops, the only question being what possible argument could justify it when the very activity of providing justificatory argument has to be inconsistent with its intended conclusion? It may be justified to believe that believing is beyond the boundaries of vervet competence, but someone who pretends that it is beyond his own is a walking counterinstance to his own scruple.

Self-counterinstantiation is relatively simple to avoid for those who practice the basic physical sciences, largely because the predicates true of their objects are not typically also true of those scientists, at least as scientists. Some logical footwork is doubtless needed to step around those references to observers and experimental intervenors, internal reference to which marks the revolutions in physics of recent times, but in the main the basic scientist may hang his self-descriptive vocabularies outside the laboratory door. In psychology, alas, as in all the so-called human sciences, the outside of the laboratory is the inside of the laboratory, and the decision to leave a vocabulary behind is typical of what we need the mooted vocabulary to describe. The danger in grounding our conception of the universe finally on the universe as described in the basic sciences is that this vocabulary lacks the wherewithal for representing science itself, which tends to forget that it is part of the world. Nothing has so impeded philosophical clarity on the topic of behavioral science as the failure to reckon that the sciences themselves are part of the reality to be dealt with. Once the behavioral scientist admits that intentionalistic descriptions fit him to a T, the question he may face, no longer one of principle but of fact, is the degree to which, only for example, vervet alarm signals at all resemble the cries of "antiscientific!" the hardliner emits at the approach of dangerous mentalists like Dennett. And Dennett has given some preliminary procedures for making a plausible judgment.

2. Dennett appears to think that intentionalist description is applicable only *grosso modo,* to be replaced in time with some finer-grained account, and the question I would raise for him is whether intentional concepts are to disappear from the latter. I am dubious about that prospect for reasons parallel to those canvased in the above comical skirmish with behaviorism. There is a form of eliminative materialism abroad these days, according to which propositional attitudes (such as "believes that," "intends that," and the like) will be replaced, together with the entire theory – folk psychology – in whose vocabulary they play a major role, with some advanced neuroscience from whose vocabulary they shall be absent (P. M. Churchland 1981; Danto 1983). Now if we are speaking of tomorrow's neuroscience, we are surely speaking of a *science,* hence of a highly intentionalized activity. Its practitioners will perform experiments, will draw inferences from the outcomes of those experiments, and on the basis of all this will arrive at some more or less warranted beliefs about ourselves (and themselves). So the terms purged from their vocabularies will still be required to characterize the fact that these are *vocabularies,* and as such play a semantic role in the theories of the sciences, themselves representations of the world according to those sciences. The materialist, like the behaviorist, is then in danger of being a walking counterinstance to his own theories. For he *possesses* a theory, a theory is *about* something, and if the theory is about states of himself that are themselves not about anything at all, where is the theory in relation to him? If aboutness disappears, so does the theory, and so does science.

So I am not at all certain that we are dealing with something transitional, which will yield in time to a finer-grained account. Of course a finer-grained account of intentionality itself will doubtless be expected – the explicit history of our philosophical consciousness of these concepts is surprisingly short, and they are under intense logical investigation today. And the manner in which intentional structures are physiologically embodied is high on our philosophical shopping list. This will be a theory that uses intentional language to mention intentional structures the users of the theory typically embody if the theory is true.

3. For these reasons I feel that intentionalism really is a theory and not a mere stance. Intentionalist stances may be taken toward whatever it amuses us to think of intentionalistically. Someone may construe twinkles as the language of the stars, and imagine stellar communications transmitted from star to star across the icy heavens. Inversely, it may please the behaviorist to think of himself as a mouse writ large. If intentionalism were only this, it would be but someone's fancy to treat vervets as scientists writ small. Perhaps science begins in stance, but at a certain moment it graduates into theory, with the supposition that it is legitimate to organize the data a certain way because that is the way they are organized *in rebus.*

I have tried to argue that to treat intentionality as a posture is to forget that in the very execution of science we have to have gone beyond posture to theory, for the taking of the posture transforms it into theory, as we exemplify it from the start. Science is not simply something we do, it is something we are; and if that is true, so is intentionalism. If it is false, the truth of intentionalism follows from our so much as believing it true.

Adaptationism was always predictive and needed no defense

Richard Dawkins

Department of Zoology, University of Oxford, Oxford OX1 3PS, England

An unkind definition of a certain type of philosopher is someone who doesn't have a subject of his own, but thinks he can do other people's subjects better than they can. My own subject of evolutionary ethology is particularly vulnerable, since we tend to write in plainer English than, say, nuclear physicists or immunologists, and any philosopher can quickly pick up enough smatterings to sound important (e.g. Clark 1982; Midgley 1979). But sometimes one comes across a philosopher who really does have something original and useful to say to biologists, possibly not *because* he is a philosopher, but simply because he is a bright guy – he might have done even better if he had gone into biology in the first place! Such a one, I suspect, is Daniel Dennett. I have never failed to be stimulated and excited by his writings, and the present target article is no exception.

His discussion of "intentional systems" will be compulsory reading for my tutorial students of animal communication. They will enjoy it once they have got used to the surprising philosophical practice of applying the concept of "intention" to a verbal meaning rather than to a muscular action. The lapwing's soliloquy – "all of a sudden I feel this tremendous urge to do that silly broken-wing dance" – is a delight, and I shall certainly steal it in lectures. I suppose my own "manipulation" approach to animal communication is most naturally seen as near the killjoy end of Dennett's spectrum, but I would like to think that it could be applied at least some of his higher levels. Animal signals don't have to "mean" anything at all: They might have more in common with spellbinding oratory, hypnosis, subliminal advertising, or direct electrical stimulation of the brain (Dawkins & Krebs 1978). Couldn't Dennett construct a similar spectrum, from killjoy upward, for a "manipulational" rather than an "informational" interpretation of animal signals? If not, perhaps this fact is important in its own right. Before leaving the first part of Dennett's paper, I think it should not be forgotten that, as Dennett himself implies, ethologists are no longer wholly wedded to "pure behaviorese." Within the ranks of ethologists the first tentative steps have been taken toward a "cognitive ethology," toward developing experimental and observational methods (M. Dawkins 1980) of tackling "the question of animal awareness" (Griffin 1981).

Those are all just passing reflections. What I really want to talk about is the "Panglossian paradigm" (the witty borrowing from Voltaire is, of course, J. B. S. Haldane's though he used it, more aptly than Gould & Lewontin 1979, to attack naive group-selectionist or nonselectionist interpretations of adaptation; J. Maynard Smith [*q. v.*], personal communication). I was infuriated by the ignorant party-line toeing of Dennett's anonymous critic who claimed that Lewontin and Gould had "shown" adaptationism to be "completely bankrupt." I urge that critic to read some of the other papers in the same Royal Society symposium that contains the overrated "Spandrels of San Marco" paper, especially Clutton-Brock and Harvey's "Comparison and Adaptation" (1979). Dennett's rebuking of his critic is, of course, entirely just, but he still concedes too much. His whole mischievous parallel with Skinner [*q. v.*] is based upon the concession that adaptive hypotheses are untestable, or at least are as difficult to test as mentalistic ones. He is right that such hypotheses could still be valuable, even if they *were* as difficult to test as mentalistic or "intentional" hypotheses. One can argue for a long time about the testability of mentalistic hypotheses and, to put it mildly, Dennett has a long way to go before he convinces most practical ethologists that they really can use the Sherlock Holmes method in the field. But hypotheses about adaptation have shown themselves in practice, over and over again, to be easily testable, by the ordinary, mundane methods of science.

Ethologists such as Tinbergen (1965) have used experimental tests in the field with great success over many years. Clutton-Brock and Harvey (1979) show the effectiveness of nonexperimental statistical methods, while at the same time destroying the myth, started by Julian Huxley and still uncritically repeated 40 years later, that allometric relationships are nonadaptive. There is, indeed, a flourishing school of comparative adaptationists, using statistical methods to test adaptive hypotheses with just the same rigour as those same statistical methods permit in any other biological context (Clutton-Brock & Harvey 1979; Eisenberg 1981; Oster & Wilson 1978; Ridley 1983). Gould and Lewontin, of course, know this perfectly well, which underlines Dennett's observation of the mismatch between their rhetoric and their conclusions. I shall discuss some cases briefly.

If log brain weight is plotted against log body weight for a variety of mammals, the points roughly fall about a straight line of ⅔ (Jerison 1973) or ¾ (Martin 1981) slope, but some points are scattered above the average line, others below, and this is interesting. The points for rodents lie around a parallel line of slightly lower intercept than the average line for mammals as a whole, but within rodents there is still some functionally interesting scatter: Forest dwellers have larger brains, body size for body size, than grassland forms; tree-climbing rodents have relatively larger brains that digging species; nocturnal rodents have relatively larger brains than diurnal ones; rodents that eat insects or fruit have relatively larger brains than those that eat leaves. All these facts suggest interesting adaptive hypotheses, but there is some danger of confounding of effects. When correlations between these various ecological dimensions for rodents are "partialed out," brain size remains correlated only with diet: Leaf eaters are relatively small-brained, and, interestingly, the same is true of primates (Clutton-Brock & Harvey 1979; Harvey, Clutton-Brock & Mace 1980). It is now our duty to think of specific hypotheses to explain this relationship, preferably hypotheses that make further predictions by which they may be differentiated.

Big primates have bigger testes than small primates, but, again, some species have relatively larger testes for their body size than others: The points on the plot of testis size against body size are scattered about the average line rather than lying precisely upon it. A specific adaptive hypothesis is that in those species in which females mate with more than one male, the males need bigger testes than in those species in which mating is monogamous or polygynous: A male whose sperms may be directly competing with the sperms of another male in the body of a female needs lots of sperms to succeed in the competition, and hence big testes. Sure enough, if the points on the testis-weight/body-weight scattergram are examined, it turns out that those above the average line are nearly all from species in which females mate with more than one male; those below the line are all from monogamous or polygynous species. The prediction from the adaptationist hypothesis could easily have been falsified. In fact it was borne out (Harcourt, Harvey, Larson & Short 1981). Notice that this hypothesis specifically links testis size with male competition *after copulation* – sperm competition. In species whose males hold harems, most male competition occurs before copulation – competition for females. We should therefore expect the males of such species to be relatively well equipped to fight. Again, using similar methods which I will not go into, the prediction is confirmed (Clutton-Brock, Harvey & Rudder 1977). Again, it could easily have been falsified.

But, it may be said, if the prediction *had* been falsified, wouldn't the adaptationist simply have changed his hypothesis, and gone on changing it until he found one that was not falsified? Well *of course* he would – that is what science is all about! It is true that the one hypothesis we shall never test is the hypothesis of no adaptive function at all, but only because that is the one hypothesis in this whole area that really *is* untestable. None of these methods, of course, tests the extreme adaptationist hypothesis that the adaptation seen is the best *conceivable*, but the man that holds that hypothesis is a straw man indeed. To the modern adaptationist, constraints on perfection are fascinating objects of study in their own right (R. Dawkins 1982, chap. 3).

Gould and Lewontin may justly be charged with precisely the same accusation they level against adaptationists: Their anti-adaptationist rhetoric (though *not* their own research papers – again the mismatch that Dennett notes) tends to suppress research by offering an easy, untestable, cop-out answer. This is most clearly seen with respect to allometry. If some bodily measure such as antler size is plotted on a log–log scale against body size across many species, and the points approximately fall around a straight line, this is *not* an excuse for saying that antler size is an automatic, nonadaptive consequence of body size (Lewontin 1979). On the contrary, both the slope and the y-intercept of the line, and the scatter of individual points about it deserve explanation. And such explanations can easily be tested by comparative examination of these measures (Clutton-Brock, Guinness & Albon 1982).

So, well done Dennett but, as far as your defense of adaptationism is concerned, you needn't have bothered. Thanks all the same. It's a lovely paper.

A la recherche du docteur Pangloss

Niles Eldredge

The American Museum of Natural History, New York, N.Y. 10024

Curious thing, this notion of "adaptation." Dennett is sensitive to a backlash against the concept in evolutionary biology – though in the one paper he consults (Gould & Lewontin 1979) he correctly notes a retreat from the initial rhetorical onslaught to a milder appeal for pluralism by the time the paper reaches its conclusion. Other comparative biologists – notably many cladistics-oriented systematists – have gone far further toward rejecting the notion of "adaptation via natural selection" as arrant vapidity than have either Gould or Lewontin who (as again Dennett correctly notes) have well-established track records as fairly conventional adaptationists.

Of course, Dennett is right when he sees that the main charge against the notion of adaptation is the ease with which anyone with a modicum of imagination and a bit of data can concoct a "just-so scenario." It has become obvious to many that in dreaming up scenarios one is merely axiomatically applying the concepts "adaptation" and "natural selection" in such a fashion and to such situations that the very testing of the validity of the explanations becomes utterly hopeless. This has indeed been a style of science (notably, if not exclusively, in paleontology, my own profession) commonly practiced for many years. It has become something of an embarrassment to the trade.

Yet the scenario-devising approach to the explanation of life's history has an honest source: The "modern synthesis" (still the dominant evolutionary theory these days) sees the modification of adaptation via natural selection as the absolutely central facet, the literal sine qua non, of evolution. All evolutionary patterns, ultrasmall (molecular), ultralarge (as in the emergence and subsequent histories of major lineages, e.g., "Mammalia"), or intermediate in size (e.g. modification of allelic frequencies within populations) are at least consistent with (and, more important, at base *effected* by) the processes collectively known as the "neo-Darwinian paradigm": Natural selection, constantly monitoring the external milieu, keeps populations adapted, perched fairly securely on their "adaptive peaks." Given enough time, it is inevitable that the environment will change.

And thus, according to this theory, it is likewise inevitable that adaptations will become modified, and new ones will emerge. Small wonder, given this view, that for so many of its practitioners the business of "doing" evolutionary biology has for so long amounted to adaptationist storytelling.

The backlash we see against this approach, it is fair to say, does not deny that organisms are adapted, that certain of their structures and behaviors can be construed as "adaptations," or that such adaptations arise through natural selection working relentlessly on a groundmass of variation. After all, few would deny that organisms seem on the whole fairly well designed and fairly well suited to the conditions of their existence – the ways they make their livings. (No one really talks about perfection these days.) At this moment, to my knowledge, the only really useful answer to the renewed creationist claim that "design in nature implies the existence of a designer" is that natural selection is a mechanistic, naturalistic, deterministic (albeit statistical) process which has a great deal of theoretical, experimental, and even "natural" ("in the wild" – though admittedly frequently "anecdotal"!) observation and analysis going for it. Not that this means that natural selection is "correct," or even that it is the only possible such naturalistic mechanism. We get back to the original bind: The complaint really is that adaptation–selection has been used too facilely, as Dennett makes clear enough (to which I can only add that its overemphasis, its firm hold on the attention of most evolutionary biologists for so many years, has tended to obscure other, equally interesting aspects of the evolutionary process, such as the prevalence of nonchange, and the differential success of species within lineages).

But I don't get the connection. I think Dennett was smeared by his earlier critics who told him his paper smacked of Pangloss. Dennett sees the connection between "adaptationism" and "intentional systems theory" as lying in the realm of optimality: Both systems seem to like to think in terms of optimality. He cites Lewontin to the effect that optimality has been the dominant reference point in evolutionary modeling in recent years, leaving the impression that the population biologists who are mainly responsible for such models are the main proponents of "adaptationism" within the ranks of evolutionary biology. But historically, right down to the present day, by far the greatest attention to adaptation has come from anatomists, systematists, and paleontologists. Let's take a brief look at the supposed research program of these sorts of "adaptationists." It can be brief because in a very real sense it doesn't exist. Consulting this literature, we characteristically find a few sentences talking of the "adaptive significance" of this or that structure or bit of behavior, or the supposed "selection forces" that produced a sequence of changes of state – an "evolutionary trend." For the most part, it's mere lip service.

But take a harder look: Beyond the papers dealing with optimal egg-clutch size, most of the "adaptationist" research in biology falls under the rubric of "functional morphology." How does that structure work; that is, how does it perform its function? Granted that a great deal of research in functional morphology has been inspired by the conviction that adaptation is what evolution is all about. (I am suggesting that rather less of such work might have been performed had the notion of adaptation enjoyed somewhat less prestige in evolutionary quarters, a blatantly untestable hypothesis!) But if you look at actual research performed by functional morphologists you will find sober analyses of fulcra, force vectors, and so forth: the understanding of anatomy as a living machine. Some of this stuff is very good. Some of it is absolutely dreadful – full of fantasies and BS (not meaning behavioral sciences in this instance). Using flume tanks and various clever tricks, Dan Fisher (1975) showed that modern horseshoe crabs, which swim on their backs, do so inclined at an angle of some 20–30 degrees, achieving a maximum speed of some 10–15 cm/sec. Assuming only that Jurassic horseshoe crabs also swam on their backs, Fisher showed they must have swum at an angle of 0–10 degrees (flat on their backs) and at the somewhat greater speed of 15–20 cm/sec. Thus the "adaptive significance" of the slight differences in anatomy between modern horseshoe crabs and their 150-million-year-old relatives is translated into an understanding of their slightly different swimming capabilities. (In all honesty, I must also report that Fisher does use optimality in his arguments: He sees the differences between the two species as a sort of trade-off, where the slightly more efficient Jurassic swimmers appear to have used the same pieces of anatomy to burrow somewhat less efficiently than their modern-day relatives.) In any case, Fisher's work stands as a really good example of functional morphological analysis. The notion of adaptation is naught but conceptual filigree – one that may have played a role in motivating the research, but one that was not vital to the research itself.

And so to "mentalism" and "intentional systems" in cognitive psychology, versus "behaviorism." From the standpoint of "adaptationism," they don't seem all that different to me, as an outsider: Both seem to share an overall goal of analyzing the function of behavior, meaning "how it works." No one, I take it, disputes the interpretation of the vervet calls as alarms occasioned by the arrival of a predator (which, of course, is an adaptation, whether or not of a sociobiological stripe). The question is, How do these calls work? Does Tom mean "martial eagle"; does Sam know Tom sees an eagle? Does "behaviorism," any less than "mentalism," seek to understand how those calls work? One approach is spare, "minimalist," and apparently disappointingly devoid of the sorts of phenomena that delight philosophers of Dennett's ilk. The other – his "intentional systems" – is arguably more romantic, but not thereby rendered a priori a less accurate view of the behavior of vervets. And Dennett, very cleverly, shows how his preferred approach can in Sherlock Holmesian fashion potentially yield reliable results – trustworthy statements of the kind Fisher produced for horseshoe crab behavior, past as well as present. Perhaps behaviorist research strategies can do as well. I simply do not know. But when we ask which approach is more Panglossian, more similar to the "adaptationism" troubling evolutionary biology of late, we really aren't asking much of a question. M. Pangloss, it turns out, has only symbolically inspired research. His is but a vague rationale for the otherwise thoroughly legitimate quest to figure out how things work. The relative merits of "behaviorism" and "mentalism" will have to be decided on other grounds. We need no longer search for the research du docteur Pangloss.

Lloyd Morgan's canon in evolutionary context

Michael T. Ghiselin
Department of Biology, University of Utah, Salt Lake City, Utah 84112

Dennett's proposals for a better approach to the study of cognitive ethology are hardly unprecedented. Sophisticated and non-Panglossian approaches to the study of adaptation have long been available, and the behavioral sciences include some good examples (Darwin 1872; Süffert 1932). Indeed, the "Panglossian paradigm" is a myth, analogous to the "anecdotal school" of psychology, an expedient motivated by academic salesmen, which continues to be discussed in the literature. In discussions of both we are rarely presented with more than slogans. What is wrong with an anecdote anyway? A respectable ethologist might call Dennett's anecdotal evidence nothing more than good observation in the field. I wonder what would happen if we tried to eliminate anecdotal evidence of this sort from volcanology.

Scientists tend to be quite vague when they invoke canons of evidence, and often apply them mechanically, without giving much thought to the underlying rationale. Perhaps this is because they are used more often at the writing desk than in the laboratory, as an expedient way of putting down the opposition. Parsimony is a case in point. Not only is it mainly invoked as a last resort; what qualifies a theory as parsimonious is not always clear. I think that Dennett, like many others, misinterprets Lloyd Morgan's canon when he treats it as an instance of parsimony. At least it is not the sort of parsimony with which I am familiar in ordinary scientific discourse. To be sure, a lot of entities are conflated under that rubric, so that there is adequate justification for a casual use of the term. Be this as it may, when I have invoked parsimony in my own work (e.g., Ghiselin 1974), I have reasoned basically as follows.

Theory A explains facts X and Y, if and only if ad hoc hypothesis B is also invoked.

Theory C explains facts X and Y, without recourse to any other hypothesis.

Therefore C is preferable to A.

This gives a *logically* simpler explanation, which is not to be confused of course with other kinds of simplicity, such as physical simplicity or ease of conception. Lloyd Morgan does seem to have had something else in mind. Indeed, he explicitly denied that his canon was a version of parsimony, which could in fact be invoked on the other side (Morgan 1909, p. 54). A higher faculty could explain more behavioral facts than a lower one. The behaviorist ad hoc hypotheses that can always save appearances might reasonably be rejected on the basis of parsimony, even though they are more in line with Lloyd Morgan's canon. Lloyd Morgan rather argued that the lower faculties will evolve before the higher ones. The lower faculties should be more common, and more widely distributed taxonomically. Hence we should prefer what is most likely. It is not a matter of logical simplicity, but of theoretical probability. The canon itself aside, it should interest us that we might use evolutionary theory as a means of evaluating cognitive and adaptative hypotheses. A true hypothesis about either must not contradict a law about evolution, any more than it may contradict any other truth. The examples Dennett gives of "rationalism" turning out to be the results of trial and error are paralleled in many recent developments in evolutionary epistemology, among the most interesting of which is the theory of common law (see Ghiselin 1982).

I have argued against Panglossianism at great length (Ghiselin 1969; 1974), and some of my arguments have been pressed into good service by Gould (1980). It was not, however, philosophical reflection in a vacuum, but the realization that a better approach was possible, that led me to discover the adaptive significance of various "reproductive strategies," including those for sequential hermaphroditism and sex itself. Much of the recent progress in sociobiology has resulted because we have rejected group selection along with its "adapted species," thanks in no small measure to a seminal book by Williams (1966). Unfortunately, the point has not been adequately appreciated, especially by those who have developed a Panglossianism of the gene to replace the Panglossianism of the soma and population. To deal with adaptation effectively, one needs a new ontology, and a new heuristic.

Panglossianism is bad because it asks the wrong question, namely, What is good? I agree with Dennett that it has some heuristic utility, for it generates (an inadequate range of) hypotheses, some of which contain a small measure of truth. Unfortunately, once a Panglossian comes up with a plausible reason, he "satisfices," and publishes a "just-so story." The alternative is to reject such teleology altogether. Instead of asking, What is good? we ask, What has happened? The new question does everything we could expect the old one to do, and a lot more besides. It generates more – and better – hypotheses, and also critical tests.

Evolutionary biology is a nomothetic science, not just a historical one. Its laws of nature generate predictions that are strictly necessary – true of everything to which they apply, irrespective of time and place. These laws govern adaptation. They tell us what sorts of adaptations can and cannot evolve under given conditions. Evolutionary theory predicts that both adaptations and maladaptations will occur, and that there will be no "prospective" or "species-level" adaptations as a result of natural selection. Laws of nature, however, are generalizations about classes of individuals. In the case of natural selection, the individuals in question are species, not organisms. Therefore any discussion of adaptation at the organismal level is meaningless apart from the context of what has happened to populations. (For a discussion of species as individuals see Ghiselin 1981.)

It is an egregious blunder to claim that the study of evolutionary adaptation posits optimality in any interesting or significant way. Evolutionary theory predicts that selection pressure under certain conditions will shift gene frequencies one way or another, just as in physics a body will respond in a certain way when acted upon by a force. A biologist who posited adaptation would be like a physicist who posited that bodies fall. Competent biologists treat the occurrence of adaptation or maladaptation as contingent in the same way that competent physicists treat the rising and falling of bodies as contingent. Adaptation has to be hypothesized and tested like everything else in science if it is not to be odiously metaphysical.

In responding to any call for pluralism we should beware of its dangers. On the one hand we must not oversimplify the world itself and ignore those diversities and complexities that really exist. On the other hand, the progress of investigation stops dead in its tracks when we allow contradictory elements to coexist within any body of knowledge, including its entirety. Monism is to be preferred in the sense that the universe is a single individual the truths about which constitute a whole. Any pluralism that leads to inconsistency and incoherence must be spurned as pernicious. That was the problem with many efforts on the part of early advocates of the synthetic theory to deal with adaptation (such as Rensch, Darlington, and Stebbins; see Ghiselin 1974). They had a fact for every hypothesis and a hypothesis for every fact. If a feature was not of advantage to the individual, it was of advantage to the species – or perhaps a pleiotropic by-product. We had more of a mixture than a synthesis. The new adaptational biology is neither Panglossian nor pluralistic, but tests broad, general hypotheses against hard data and is not satisfied until all contradictions have been purged from the system (e.g. Charnov 1982). This paradigm, however, is Darwin's paradigm, revived and modernized. Evolutionary biology is discovering its metaphysics, and it is about time.

Denoting and demoting intentional systems

George Graham

Department of Philosophy, University of Alabama in Birmingham, Birmingham, Ala. 35294

Do we know what it means to believe and desire? Well, we know, I think, what we want ascriptions of desire to do, for example. As Dennett points out, when we ascribe a desire to an animal, we want to explain why it behaves as it does. In particular, I would argue, we want to include reference to the consequences of its behavior in the explanation of the behavior. "Why does the monkey yell?" "Because it wants to ward off predators." Warding off predators figures as the object of the monkey's desire, even though it is an event that (at best) may occur. Through the mediation of the mental state (the desire), it becomes a factor in a current state of affairs. It gets "represented" in the monkey's head.

Behaviorists urge skepticism about ascriptions of beliefs and desires. For one thing, such ascriptions are too tempting, too easy. And if behaviorists are right, there are experimentally more tractable explanations in terms of past consequences of behavior. For another, such ascriptions raise difficult metaphysical questions. How can something that is only possible be represented in the head? What separates one possible state of affairs from another?

Dennett shares some behaviorist skepticism, but wishes ethologists to appreciate the utility of ascriptions of beliefs and desires anyway. It all comes down to this: Explanations in terms of beliefs and desires can be useful, even if they are not true. And, for Dennett, they are not true.

How can they be useful? By guiding explanations that are true; explanations that avoid (transcend, overcome) attributing beliefs and desires.

Why are only explanations that avoid attributing beliefs and desires true?

It is not made clear in the target article why Dennett thinks that only explanations that avoid ascribing beliefs and desires are true; why attributing beliefs and desires is "interim" speech until animals can be "demoted" or redescribed as "dumb" physical systems. Granted that intentional state ascriptions are too tempting, too easy, and raise hard metaphysical questions, they might still be factually correct. Although, if they are factually correct, we need a distinction that marks when they are. This is a distinction that Dennett does not make.

I have in mind a distinction between *real* and *nominal* intentional systems. The idea would be that certain animals believe and desire; others do not. Those animals that believe and desire are real intentional systems. Animals that do not believe and desire but that are usefully thought of as doing so are nominal intentional systems.

Some ethologists should accept the above distinction even if Dennett does not. For instance, N. K. Humphrey (1980) has tried to show – by adaptionist argument – that some social animals have beliefs and desires as well as introspective access to those beliefs and desires. Now creatures could not introspect beliefs and desires without having beliefs and desires. So, if Humphrey is right, and certain animals introspect beliefs and desires, they must be real intentional systems.

But Dennett does not make the distinction. I assume that this is because in addition to (or partly because of) the skepticism mentioned above, he believes such a distinction lacks empirical content, that is, that we cannot determine whether or what an animal in fact believes. I find such an attitude perplexing. In the first place, the distinction is arguably part of ordinary speech. It expresses our sense that some things (persons) believe whereas others (nations) do not, although we can suppose usefully that they do. Second, Dennett's intentional system theory can easily be converted into a theory that states how to test whether an animal is a real intentional system. Simply take the tests (of minimal rationality, responsiveness to novelty, and the like) that Dennett proposes as tests of the fecundity of intentional state ascriptions and treat them as tests for their truth (instead, or as well). Simply treat them as determining whether claims like "X believes that p" are true and not just otherwise apt metaphors or momentarily necessary fictions.

Of course, the main reason Dennett gives for thinking that ascriptions of intentional states lack empirical content is that they are unfalsifiable *in principle;* that revisions can always be made ad lib in order to preserve rationality. But intentional state ascriptions aren't unfalsifiable *in principle*, and making ad lib revisions is incompatible with intelligent mentalism. Intentional state ascriptions can be defeated by more systematic, experimentally sensitive (intentional or nonintentional) explanations of behavior, even if they cannot be destroyed by critical experiments. That's a kind of falsifiability, and the task facing the reflective mentalist is to clarify it.

Moreover, when I examine Dennett's discussion of the alleged unfalsifiability of intentional state ascriptions, I have the feeling that I have seen this argument before; that Dennett has made a similar point against operant behaviorism (1978a, chaps. 4, 5). He has charged that revisions can always be made ad lib to preserve reinforcement; that ascriptions of reinforcement lack empirical content because they are unfalsifiable *in principle*. But, again, they aren't. In theory, ascriptions of reinforcement can be trumped by more systematic, empirically tractable (operant or nonoperant) explanations of behavior, even when they cannot be destroyed by crucial tests. Again, that's a kind of falsifiability, and the chore facing the reflective behaviorist is to clarify it (see Schwartz & Lacey 1982 for a discussion).

Dennett has shown the potential for attributions of beliefs and desires to animals. But he stops short of saying when they are true, or distinguishing real from nominal intentional systems. Given the importance of the charge of unfalsifiability in his thinking, it is at least unfortunate that he did not explain it more clearly.

Thinking about animal thoughts

Donald R. Griffin

The Rockefeller University, New York, N.Y. 10021

Dennett's pithy analyses of some of the ways in which we can fruitfully organize our own thinking about whatever conscious thoughts and subjective feelings nonhuman animals may experience are important contributions to the early embryology of cognitive ethology. His suggestion that ethologists employ what he calls "intentional system theory" as a sort of working hypothesis or theoretical framework to suggest improved observations and experiments may very well encourage many students of animal behavior to gather more significant data in their future observations and experiments. I hope so, but at the same time I suspect that this is only a first step.

Why be so timid? Why not face up explicitly to the possible reality of animal thoughts and feelings? Why not call a thought a thought, an emotion an emotion, and so forth? Conscious, subjective experiences of animals may indeed be real and important to them and to our understanding of them. "Stances," working hypotheses, and models are all very well as interim measures, but the major questions of cognitive ethology concern whatever conscious thoughts and subjective feelings actually do occur in nonhuman animals. To urge that we slide into cognitive ethology by a side door, so to speak, tends to perpetuate the behavioristic myth that conscious experience is unimportant. Dennett's suggested "intentional stance" may help some scientists to complete their escape from the behavioristic instar. But do we still need such a positivistic security blanket?

Conscious experiences are notoriously elusive phenomena, but this certainly does not mean that they are insignificant in people, and perhaps not in other animals either. The serious scientific problems involve how to study them, and recognizing their possible reality and importance is a useful, indeed almost certainly a necessary first step.

I have suggested elsewhere (Griffin 1978; 1981) that communicative behavior may serve to convey thoughts and feelings from one animal to another. If so, the interception and accurate interpretation of communicated messages can tell us at least part of what the communicator is thinking or feeling. Others may prefer different approaches, but the important task is to get on with the job of developing improved methods for detecting and analyzing conscious experience in other species. The early stages of such investigations may well be incomplete and imperfect; but that has been true of almost every scientific development. Hence the prospect of an uncertain ontogeny does not justify discarding the cognitive baby with its behavioristic bathwater.

Adaptationist theorizing and intentional system theory

Gilbert Harman
Department of Philosophy, Princeton University, Princeton, N.J. 08544

1. What do adaptationist theorizing and intentional system theorizing have in common? Answer: They offer functional explanations – in both cases S's doing D is explained by noting that there is a certain function or purpose or rationale for S to do D. Intentional system theorizing offers a mentalistic explanation that cites S's *awareness* that doing D has a certain function. Adaptationist theorizing appeals to what Nozick (1974) has called an "invisible hand" explanation. For example, an adaptationist explanation of S's doing D might be that the possession of a genetic makeup that leads its possessor to do D is why organisms with those genes have survived.

Dennett notes the resemblance between recent antiadaptationist arguments and behavioristic arguments against mentalism. I am struck also by the similarity between adaptationist theorizing and Skinner's behavioristic theory of operant conditioning. Operant conditioning too offers a kind of "invisible hand" explanation, namely that behavior that functions well tends to be rewarded, and therefore survives, compared with other behavior. And the objections Dennett cites against adaptationist theorizing resemble familiar objections to the theory of operant conditioning – it is too easy to produce explanations of the relevant form, the theory is unfalsifiable, the approach distracts attention from serious empirical questions, and so on.

In particular, if mentalistic explanations and adaptationist explanations are in principle unfalsifiable, this is true in exactly the same respect of explanations in terms of operant conditioning. But then, it is true of *any* serious explanation in science! We cannot ignore the familiar Duhemian point that no theory can be tested by itself (Duhem 1906; Hempel 1950; Quine 1951). Auxiliary assumptions of various sorts are always needed – for example, assumptions about the measuring apparatus. This means that there can be no logically crucial experimental test of the theory; any failure of testing can always be blamed on auxiliary assumptions rather than on principles of the favored theory.

Dennett argues that adaptationism and mentalism should not be taken to be theories but rather represent what he calls "stances or strategies that serve to organize data, explain interrelations, and generate questions to ask Nature. Were they theories in the 'classical' mold," he says, "the objection that they are question begging or irrefutable would be fatal, but to make this objection is to misread their point." But since any theory can always be made question begging and irrefutable, the point does not distinguish adaptationism, mentalism, and operant conditioning from theories in the "classical" mold. In this sense, all theories are "stances or strategies."

2. Dennett says that mentalistic vocabulary can be picked out by testing for "referential opacity." But this test also picks out some not obviously mentalistic vocabulary, such as "it is provable that," "it ought to be that," "it is necessary that," "because." To illustrate the point, consider that a stone initially at rest at a point midway between the sun and Mercury will be pulled toward the sun. That is because the sun is the more massive of the two bodies. It is not because the sun is the sun, or because the sun is the brighter of the two bodies, even though the more massive of the two bodies = the brighter of the two bodies. Substitutivity of identity fails in a context of explanation; such a context is referentially opaque. But there is no obviously mental terminology involved.

I am not sure whether this matters, however, since I am not sure why Dennett mentions referential opacity or how it is relevant to anything else in his article.

3. Dennett notes Grice's (1957) view that to say someone is communicating is to ascribe a third-order intentional state to the person, at a minimum. Actually, in Grice (1957) there is some risk that the level will be infinite, not just third order. His view is that Utterer must intend Audience to produce response r in virtue of Audience's recognition of what Utterer intends. This avoids an infinite regress only if the relevant intention is in part about itself! Utterer intends Audience to produce response r in virtue of Audience's recognition of *this very intention* (Harman 1974). It is not clear what "level" this sort of self-referential intention would be on. Perhaps it is on Dennett's second level.

4. The devious vervet, who issued a leopard alarm in the absence of any leopards while his side was losing a monkey fight, may have relied only on its relatively low-level knowledge that monkeys head for trees when that sound is made, even if it also happens to have a higher-level view about why monkeys head for trees when the sound is made.

5. One way to test whether a bird that does the "broken-wing dance" does it because of its reasons to do it would be to see whether it does the "dance" only when there are chicks in the nest.

Belief ascription, parsimony, and rationality

John Heil
Department of Psychology, University of California, Berkeley, Calif. 94720

Dennett's defense of the "Panglossian paradigm" seems to me philosophically sound. The comments that follow are addressed to three related matters. They are offered not as criticisms of particular doctrines, but merely as points of discussion.

1. What, one may wonder, is the connection between a creature's having certain "higher-order" intentional states – states the content of which includes some further intentional component – and that creature's employing a language? Dennett argues (following Grice 1957; 1969) that before behavior can be regarded as "genuinely communicative" it must be appropriately linked to such higher-order intentional states. It is not enough, of course, that a creature merely possess a complement of such states. These must, in addition, play a certain role in the production of utterances. It seems natural to imagine, then, that an investigator concerned with the linguistic prowess of some creature must first ascertain the character and content of that creature's psychological states, then bring this information to bear in a determination of the semantic content of the creature's "utterances." Once an investigator assigns a particular content to an utterance, he thereby imputes a certain higher-order content to the utterer's states of mind.

An apparent snag in such a project, however, stems from the fact that much of one's evidence for the presence of higher-order intentional states seems to hinge on one's prior identification of behavior as linguistic. This is perhaps an unexpected consequence of Morgan's (1894) principle of psychological parsimony (mentioned by Dennett; see below). On the latter principle, in the course of ascribing psychological states to creatures, one is bound to advance only the most parsimonious interpretation consistent with the behavioral data. One is entitled to move to a higher level of intentional state ascription only if one's evidence, in a certain sense, forces one to do so. Granted, any intentional interpretation will be underdetermined. Still, one seems obliged to minimize this underdetermination at least to the extent of opting for a simpler, lower-level ascription given the option.

Now what sorts of behavior would in this way *force* one to ascribe higher-order intentional states to some creature? Suppose that a creature, S, sets a trap for *T* by digging a pit and camouflaging it. Here we shall doubtless be obliged to ascribe a host of beliefs to S. But should we also ascribe to S beliefs about, for instance, *T*'s states of mind, more particularly, beliefs about

T's beliefs? We *can* do so, of course, but would we be entitled to do so? We may suppose, for instance, that *S* believes that *T*, on encountering the camouflaged pit, will believe that his path is safe, or perhaps that *T* will not come to believe that there is a pit in front of him. But it may be far simpler, given the circumstances, to suppose only that S believes that in constructing a pit in this way he can ensnare *T*. Such a belief does not – not obviously, anyway – require the ascription to S of any higher-order beliefs.

I suspect, although I cannot argue the matter here (see Heil 1982; 1983), that evidence for the possession of higher-order intentional states must be linguistic in character (and linguistic in some fairly strong sense). I do not mean that one must first set out to identify behavior (somehow) as linguistic and then, on the basis of this identification, proceed to the ascription of higher-order beliefs, desires, and intentions. Such a procedure would fail for reasons roughly parallel to the reasons for which attempts to identify instances of higher-order intentional states without recourse to utterances fail. The only plausible evidence we can have for the presence of the one is of a piece with the evidence we can have for the other.

If these remarks are not altogether mistaken, then I suspect that it is at the very least misleading to assimilate language to communication (even "genuine communication"). Communication, in the ordinary sense, seems not to require any special sort of intentional backing. I may communicate to you the fact that I have a cold by coughing in my sleep. The clicking sound issuing from beneath the hood of your DeSoto provides eloquent testimony that a valve adjustment is called for. In this sense, it seems uncontroversial to suppose that nonhuman creatures communicate all sorts of things to other, sufficiently observant, nonhuman creatures. Even where such communication results from what appears to be voluntary action, however, we should be cautious about describing it as linguistic – for in so doing we also ascribe an assortment of higher-order intentional states to the creatures involved.

2. One may suspect that the principle of parsimony mentioned above differs importantly from the familiar notion of simplicity one associates with theory construction in the natural sciences. Thus, whether or not a given intentional state attribution is more parsimonious than another is not something one can ascertain merely by comparing the two attributions.

We encounter S skating on a frozen pond and attribute to him the belief that the ice is thick enough to support his weight. We might, however, wish to attribute to him an apparently more complex belief – for instance, the belief that the ice is in fact quite thin, but that it will continue to support his weight so long as he has faith (continues to believe) that it will.

Whether our ascription of the latter belief is indeed less parsimonious than an attribution of the former is not something one can decide merely by comparing the two beliefs – even if one notices that the second belief is at a higher intentional level than the first. Rather, we should regard one belief ascription as more parsimonious than another only if it fits better with our "total theory" of S's intentional states and processes. Thus, if we suppose S to hold certain religious beliefs, for example, and beliefs about the ice on this pond, the ascription to him of the second of the two beliefs mentioned may turn out to be, for us (and given our overall interpretation of S), more parsimonious – that is, its ascription would not oblige us to revise our theory in ways that are not independently motivated.

If this is right, then the requirement of parsimony in the ascription of psychological states may perhaps be regarded as a corollary to the more general requirement of rationality.

3. Finally, the principle articulated by Dennett in (14) does not seem to me to be even partially constitutive of rationality. It may, that is, be too strong to claim that a rational agent who holds p must also believe q when p implies q. *S* may hold p and hold as well (and with good reason) not-q. *S* may, for example, not *recognize* that p implies q. Even if *S* does recognize this implication, however, rationality requres only that *S* ought *either* to hold q or abandon his belief in p.

In any case, it will not, in general, do to ascribe beliefs to agents whom we suppose to be rational by applying such principles as (14). It is surely too strong to say that such agents must believe whatever is entailed by what they hold true – must believe, that is, all the consequences of their beliefs. One may, it seems, justifiably hold p and not-q, even when p implies q. So long as one does not hold the second-order belief that p implies q, one's rationality in such cases need not be impugned.

The *adaptiveness* of *mentalism*?

Nicholas Humphrey
Sub-Department of Animal Behaviour, University of Cambridge, Madingley, Cambridge CB3 8AA, England

Excited as I am by Dennett's paper, I remain puzzled about the relation between the first and second halves. Why, *in the context of this paper*, does Dennett go so far out of his way to defend the "Panglossian paradigm"? What has the usefulness – or lack of it – of the intentional stance really got to do with the question of whether we do – or do not – live in the best of all possible worlds?

At first reading anyway, the discussion in the second part struck me – to use Dennett's own example – as a kind of distraction display: "All of a sudden I feel this tremendous urge to do that have-a-go-at-Gould-Lewontin-and-Skinner routine. I wonder why?" Well I wondered why. And I would offer Dennett a more solid excuse than he himself, in his discussion of "rationality," provides.

Dennett's line is this. Cognitive psychologists who, as "mentalists," assume that other living beings have intentional states, are ideologically in the same boat as evolutionary biologists who, as "adaptationists," assume that everything they find in nature is adaptive. In other words, there is an analogy between the intentional stance and what might be called the optimality stance. Buy why does he stop there? If, like me, you are already both a mentalist and an adaptationist, it seems natural to conclude that *mentalism itself* is *adaptive*. In other words, the intentional stance is itself an optimal stance – the best way of looking at behaviour that nature (or science) could possibly devise.

Thus it is no accident that Dennett – and all the rest of us "as fascinated human beings" – find the intentional stance so congenial. For human beings *have been shaped by natural selection* to think about behaviour in this optimal mentalistic way. Indeed we are, I believe, one and all "natural psychologists," predisposed by nature to use the *conscious* experience of our own intentional states as a basis for modelling the behaviour of other living beings (Humphrey 1979; 1982).

In his last paragraph Dennett claims, quite rightly, that the intentional stance "can be an immensely fruitful strategy in science." But it is much more than that. It is, for the natural psychologist, an immensely fruitful strategy in life – and a strategy that now lies deeply embedded in the human mind.

Dennett's "Panglossian paradigm"

Alison Jolly
The Rockefeller University, New York, N.Y. 10021

What Dennett's paper needs is the book that invisibly surrounds it. Some of this book has already been written by Dennett in other places. As it is, the paper does grasshopper leaps from referential opacity, to vervet Tom's speculations on the mental processes of vervet Sam, to adaptation in evolution. Various commentators no doubt mistrust these leaps, fearing that the grasshopper has blithely sailed over some chasm of illogic, while the commentator falls in.

As it happens, my naive beliefs largely coincide with Dennett's more elegantly explained concepts. Therefore, let me work backward from his conclusions to a few points where I wish he would write in more of the bridges.

Dennett concludes that both adaptationism and mentalism *work*. "We are actually pretty good at picking the right constraints, the right belief and desire attributions. . . . [Adaptationism] turns out to be predictive in the real world." So does mentalism. In fact, that is why mentalism is adaptive: It is an immensely powerful way of predicting others' behavior using a few simplifying ideas like love, hate, fear, and suspicion rather than the paraphernalia of physical cause and effect. As Premack and Woodruff (1978, p. 526) remark: "It would waste the behaviorist's time to recommend parsimony to an ape. The ape could only be a mentalist. Unless we are badly mistaken, he is not intelligent enough to be a behaviorist."

One of the most revealing phrases of the target article is the computer's similar slip: "Gravity drowned." This is far more than prowess mixed with stupidity; this is the computer assuming that any agent is alive. Of course, the human author of the program slipped in not distinguishing animate from inanimate agents, but this is a "usually reliable shortcut." This same shortcut may have infused our primate ancestor's diffuse conceptions of external cause and effect (the lion wants to eat me, the river wants to drown me) as well as far later human attempts to find logic in the external world. It would be pleasant to convert TALESPIN into MYTHSPIN by changing a few names, such as turning "gravity" into "servant of Mother Earth."

Suppose imputing intention is, in fact, very simple (and simplifying) and quite ancient in the primate line, or even common to many social mammals, not just humans. Do Dennett's earlier sections clarify our thoughts on this? Somewhat, but I wish Dennett had dealt with the difference between desires and beliefs. Desires intuitively seem much simpler than beliefs. Further, the game of referential opacity seems to work only when the intentional state ends with a noun clause; that is, is a belief not just an emotion, as with "Burgess fears *the rustling in the bush is a python*." Suppose Burgess is a simple-minded elephant shrew, or is a human in the first second of a terrified startle response. We could only say "Burgess fears *the rustling in the bush*." There is nothing to substitute, unless we want to add that in the course of evolution or experience such rustling implies probable danger, and therefore Burgess fears *probable danger*, which is both true and has survival value.

Imputations of emotion to others may also be simpler than imputations of belief. We then have

1.0 order: Tom wants Sam to run into the trees
1.5 order: Tom wants Sam to feel leopard-flavored fear (which may lead Sam to a better choice of trees and escape routes, though Tom needn't calculate this)
2.0 order: Tom wants Sam to believe there is a leopard

It is difficult to design experiments that differentiate between these three versions. Even if there *isn't* a leopard – as Dennett mentions, Seyfarth and Cheney once heard a vervet break up a territorial battle which his side was losing by giving a leopard bark – any of the three versions above could still apply. Premack is now testing chimpanzees' attributions of beliefs and desires to their trainers; he feels that attributing desires may be much easier for a chimp than attributing states of knowledge. We have, however, a few anecdotes of attributed states of knowledge, as in the female hamadryas baboon who spent 20 minutes inching forward to hide her forequarters behind a rock, where she groomed a forbidden subadult male. She left her back innocently in view of the harem overlord (Kummer 1982).

It seems that the problem of intention is still not cracked. It needs more of Dennett's philosophy, not less. Mentalism, as shown by chimpanzees, devious female baboons, and computers programmed with shortcuts, may well be an eminently "satisficing" means of making predictions of others' behavior. Mentalism is not so accurate, perhaps, as descending to the level of behaviorism, neuron circuitry, or nuclear physics, but it's a whole lot easier on the mind.

Elementary errors about evolution

Richard C. Lewontin
Museum of Comparative Zoology, Harvard University, Cambridge, Mass. 02138

I must confess to being both mystified and a little put off by Daniel Dennett's effort to use me both alone and together with S. J. Gould as a stick with which to beat B. F. Skinner. He begins by saying that the paper of Gould and Lewontin (1979) on adaptation was misrepresented to him as destroying the entire notion of evolutionary adaptation, whereas he found on reading the paper itself a "pluralistic" and "entirely reasonable" argument against the unthinking misuse of the concept of adaptation. Then why has Dennett gone on to write another 10 pages or so of argument and polemic on the subject of our views? If he is arguing against the view that direct natural selection for traits is never, or virtually never, the explanation of evolutionary events, then he is arguing against a position that no evolutionary biologist, least of all I, comes remotely near espousing. Indeed, were he to take a poll of evolutionary geneticists, he would find that, along the spectrum from those who favor nonselective explanations of molecular evolution to those who see selection as the dominant cause of molecular change, I am generally regarded as nearer the "selectionist" than the "neutralist" end. If, on the other hand, he is arguing against the position that nonadaptive and even counteradaptive forces play significant role in evolution, then he is simply ignorant of the developments of the last 75 years in evolutionary and population genetics. Sometimes he does one and sometimes the other, but in both cases he is wrong.

In order that the reader who is not familiar with the Gould and Lewontin paper and my other writings on adaptation not be totally mystified by the discussion, let me briefly review the issue in a way that, apparently, Dennett did not disagree with. There has been a growing tendency in evolutionary biology to reconstruct or predict evolutionary events by *assuming* that all characters are established in evolution by direct natural selection of the most adapted state, that is, the state that is an optimum "solution" to a "problem" posed by the enviornment. This is not taken as a hypothesis to be tested and possibly rejected, but as an a priori assumption, in which case the goal of theory and experiment becomes a demonstration or prediction of the *way* in which this optimal state has been or will be reached. This view of the evolutionary process, however, ignores the body of knowledge built up by evolutionary genetics, a body that, in itself, is challenged by no one. It is simply sidestepped by Panglossian adaptationists who find it inconvenient. Two examples of nonadaptive forces will suffice in the current context to make the point. First, genes are organized on chromosomes and so are transmitted together in inheritance in a correlated fashion rather than independently. As a consequence, if natural selection causes a given gene to increase in frequency and spread through a species, other genes near it on the chromosome, of no selective advantage or even of some selective *disadvantage*, can be carried along and spread along with the selected genes. This is known as "genetic hitchhiking," and its quantitative theory is extremely well worked out. Second, because of the stochastic nature of the Mendelian mechanism and because of the finiteness of actual population sizes (often they are *quite* small), genes are fixed in a species by purely random events (so-called random genetic drift) when they have no selective advantage or even when they are *deleterious in comparison to other genes already present in the population.* Indeed, a new mutation that has a selective advantage of $s\%$ over the type characterizing the population has only a chance of $2s\%$ of ever spreading through the population. Most new advantageous mutations are lost by chance.

Given these two phenomena (and others discussed by Gould and Lewontin together and separately), it is simply factually incorrect to describe evolution as always being an adaptive or optimizing process. Moreover, no use of words as they are commonly understood in English will allow those processes to be described as optimizing subject to a constraint, like the jury-rigged mast of a storm-damaged ship, as Dennett's analogy has it. To say that a population that has, for reasons of random drift or genetic hitchhiking, replaced a favorable gene by a deleterious one has thereby undergone a process of optimization on a set of constraints is ludicrous. Nature does not do "the best she can." This is not the best of all possible worlds, unless one means by "possible" exactly what has happened historically. But in that case "best" loses its meaning since only the actual was possible.

The most that Dennett is able to find to criticize in us is that the mildness of our pluralistic conclusion about evolution is overwhelmed by the rhetorical fire of our attack on adaptationism. But Dennett has confused "adaptationism" with "adaptation." We scorn the former, not the latter. We abuse a world view that raises a phenomenon to untested universality, not the phenomenon itself, which is of undoubted importance in evolution.

Unable to convict us of error on our actual words, Dennett resorts to rhetorical flummery to suggest that we really have a hidden agenda, the extirpation root and branch of adaptation, despite what we say to the contrary. What does Dennett mean when he says that "Gould and Lewontin *seem* to be championing" a "scrupulously nonadaptationist, historic-architectural answer" (his italics)? Does he have some evidence that things are not as they seem, or is this, like the infamous list of never-to-be-revealed names, a blank sheet to be waved in the air for effect? Dennett never answers his own rhetorical questions, "Could it be that Gould and Lewontin have written the latest chapter of Postpositivist Harvard Conservatism?" because, of course, the answer is no; but if he gave it, his entire attack on us as fellow travelers of Skinner would be vitiated. Could it be that Dennett is trying to implant by innuendo an idea he cannot support by logic?

Having failed to make a logical case that our pluralistic attack on extreme adaptationism is *like* Skinner's total attack on mentalism of any form or degree, Dennett actually tries to suggest that my own view of neutralism itself is Skinnerian. To this end he quotes in note 13 my remarks in the *New York Review of Books* about the current state of studies in artificial intelligence. Once again, Dennett stands the situation on its head. My criticisms of Cartesian reductionist cognitive science are precisely that the reductionist strategy has been unable to include the mental, whereas to succeed it must do so. Skinner tries to save vulgar reductionism by rejecting the mental. I reject vulgar reductionism because it fails to cope with the mental. Even the most confused philosopher should have some difficulty in reconciling those two positions.

I return to my opening remark. I am mystified by Dennett's target article. He is, I am reliably informed, among the ablest of current philosophers, yet he seems to confuse statements of the form "X is not always the case" with "X is always not the case," an error we do not allow to first-year philosophy students. Most of the points of fact and logic that I make in this commentary were communicated to Dennett when he sent me an earlier version of his paper. He answered saying that he would think them over and reply to them when he had revised his piece. He never did so. Had he, the editor, readers, he, and I might have been saved both time and expense. Biology has many difficult epistemological problems, and biologists need help with them. There is a growing list of philosophers of science, such as Philip Kitcher, Elliott Sober, and William Wimsatt, to name just three, who are making substantial contributions to biology. The addition of Daniel Dennett's acknowledged talents would be a great asset.

The scope and ingenuity of evolutionary systems

Dan Lloyd

Department of Philosophy, University of California, Santa Barbara, Calif. 93106

Since its original statement in *Content and Consciousness* (1969), Professor Dennett's philosophy of mind has been propagated, and has flourished, in some surprising environments. Though Dennett's work is largely presented in journal papers, it is becoming clear that he is at work on a systematic philosophical picture of the world, one that attempts to balance our intuitions about ourselves and our psychology with physicalist ambitions – a difficult compromise to be sure. At the heart of the world view in most of its statements is the idea of the intentional system, something to which we ascribe rehabilitated versions of our ordinary mental concepts. The central reason for making such ascriptions (taking the "intentional stance"), and so declaring something to be an intentional system, is found in the benefit won thereby: increased ability to predict the behavior of the system.

We are obvious intentional systems, but Dennett advises taking the intentional stance toward all sorts of things: from chess-playing computers (Dennett 1971) all the way to clams, thermostats, and lightning bolts (which clam up when they believe they are in danger, turn on the heat when they perceive that it is cold, and strike what they suspect offers the quickest route to ground, respectively; Dennett 1981c). We are familiar enough with the *metaphorical* extension of intentional idioms, but Dennett is advocating the literal application of such idioms in *all* the cases just listed. Humans are remote from thermostats, but in the end the cognitive difference between them is one of degree, not of kind. Dennett defends this liberalism of intentional attribution in part by stripping the concepts of belief and the like of many of their traditional mentalistic trappings: Believers, human or not, do not necessarily have infallible conscious access to beliefs (Dennett 1969; 1978b, 1979), nor, more important to cognitive ethology, need believers contain representations of what is believed (Dennett 1975). The result is an emerging picture of mind in which rationales are *always* free floating. We ascribe beliefs – to ourselves as to other entities – exclusively at our explanatory and predictive convenience.

Vervet monkeys, then, sit rather high in Dennett's great chain of intentional being. Even the birds and the bees qualify for the intentional stance. But one may have difficulty accepting intentional accounts of adaptation. The main problem is one of ascription: To what (or whom) do we ascribe the rational beliefs and desires that explain evolution? Dennett's three answers, "Mother Nature herself," "nobody," and "the evolving genotype," are all problematic. The first two are respectively too unwieldy and, so to speak, not wieldy enough: About neither can we make belief or desire ascriptions that permit prediction of individual adaptations. The use of intentional idioms to characterize genes is a familiar heuristic device in R. Dawkins (1976), but are we entitled to it in Dennett's more literal sense?

Should we say, for example, "Being sensible, genes prefer the safe course of replicating – but once in a while they exercise their mad urge to mutate." In this case the intentional stance affords little predictive perspicuity, for the reasons that it cannot predict when a mutation will occur, nor can it help us predict *which* of all possible mutations will occur. Of course, in time selection sees to it that only the wise genes survive, but the point remains that even as we look at these "smart" survivors we still have no inkling about their next move – it may be smart, but it may not be. Since genetic processes are stochastic, not rational, it seems that the intentional stance is of little value at this level. The same problem arises when we look at individual phenotypes, whose offspring will not necessarily be the rational choice for future survival.

What, then, are the intentional systems in evolution? If not individual entities, then are they *populations* of individual entities – say species? Problems of ascription remain: Suppose species A gives rise to new species B, after which A becomes extinct. Which species is credited with the wisdom of the choice to speciate? The entities that are candidates for the intentional stance in evolution thus seem to be *transspecific,* where that term directs us to look beyond both individuals and species. How large and how enduring must these systems be? [See also Ghiselin: "Categories, Life and Thinking" *BBS* 4(2) 1981.] This is an empirical question, just as is the decision to take the intentional stance in the first place. At some point the details of macroevolution will begin to make sense in the intentional framework, and the boundaries of these grand intentional systems will emerge. We might look at whole orders, even classes, as single intentional systems. The shifting gene pools of the order would constitute its mental life, individual genotypes being something like ideas. An order would perceive the world through selection, which would confirm its "good ideas" through eliminating its false starts, thus establishing species, the order's analogue to behavior.

Even so, it may still turn out to be impossible to intentionally predict the evolutionary system's next move. Would that defeat the intentional stance? Not if we remember that we can fail with our intentional predictions in either of two ways: (i) if the system in question is random (i.e., "irrational"), or (ii) if the system in question is *too ingenious* for us. If evolutionary change is unpredictable, then I would hazard that it is thus for the latter reason. Like all the actions of genius, in retrospect speciation makes perfect sense, but we could not have had the insight to identify its solutions in advance. Rather than abandon the intentional stance toward an ingenious system, we should emulate the ingenuity of the system and work toward a theory that captures it in all its glory.

Intentional system theory invites us to take this anthropomorphism literally, with the substantial fringe benefit that while we revise our conceptions of the entities that win our intentional embrace, we also revise our conception of the most familiar intentional systems – ourselves. Our essential kinship with vervet monkeys and evolutionary systems alters our view of mind and the place of persons in the physical world. Dennett's work is provocative in the best sense of the word.

Intentions as goals

David McFarland

Animal Behaviour Research Group, Department of Zoology, University of Oxford, Oxford OX1 3BJ, England

Some years ago, while engaged in a tutorial with students of animal behaviour, I observed a lone pigeon pecking at a slice of bread on the paved area outside our window. As soon as other pigeons began to appear, the pigeon quickly pecked a hole in the middle of the slice of bread, stepped on the edge so that the bread tipped up, put its head through the hole, and flew off with the bread around its neck. Not wishing to be a killjoy, I refused to deny that the pigeon had behaved intentionally.

Dennett is right in thinking that Lloyd Morgan's canon has had an undue influence upon the way behavioural science is conducted. Lloyd Morgan (1894) stated that "in no case may we interpret an action as the outcome of the exercise of a higher mental faculty, if it can be interpreted as the outcome of one which stands lower in the psychological scale." However, realising that this was being interpreted by some, later to be called the behaviourists, to deny the possibility of any but the lowest level of analysis, Morgan (1900) felt obliged to add the following rider to his canon: "To this it may be added – lest the range of the principle be misunderstood – that the canon by no means excludes the interpretation of a particular act as the outcome of the higher mental processes if we already have independent evidence of their occurrence in the agent." We are now faced with the question of what would count as independent evidence.

Dennett contrasts two general types of explanation of animal behaviour, the killjoy behaviouristic explanation and the fin de siècle (see Romanes 1882) intentional explanation. Dennett recognises that the behaviourists can usually find a low-level explanation for any particular set of observations, but he maintains that such explanations are not satisfying. "The claim that *in principle* a lowest-order story can always be told of any animal behavior . . . is no longer interesting."

Dennett shows us how intentional systems can be recognised by their properties of substitutability and rationality. He does not, however, give us the essence of intentionality. An intentional system, in essence, contains a representation of the goal (or want) that is in some way instrumental in controlling the behaviour of the animal. (It is conceivable that an animal could have a representation of the goal, or endpoint, or function, of its behaviour that did not actually influence the behaviour. Such a system would surely not be an intentional system.) In most of Dennett's examples it is implied that the representation does, in fact, control the behaviour. Thus the vervet monkey "Tom *wants* to cause Sam to run into the trees. . . On this reading the leopard cry belongs in the same general category with coming up behind someone and saying 'Boo!' Not only does its intended effect." I take this to mean that Tom has the intention of scaring Sam into the trees in the sense that this goal is represented in Tom's mind before he gives the alarm.

It is important to recognise that intentional systems involve representations that cause things to happen, because this structure is very similar to the structure of simple goal-seeking systems about which students of animal behaviour have been debating for a long time. An intention is a sophisticated version of what ethologists have long referred to as the Sollwert (Hinde 1970), efference copy (von Holst & Mittelstaedt 1950), set point (McFarland 1971; Toates 1980), and the like. These terms all refer to an internal representation that acts as a standard of comparison for feedback from the consequences of behaviour (McFarland 1971). As a result of the comparison, some action is taken. Thus Tom, seeing that Sam has not fled into the trees, gives the alarm call. When Sam has fled into the trees Tom does not give the alarm call.

The problem is that, even at the most simple level, there is always an alternative to this type of system. Technically the problem is how to distinguish between a passive and an active control system, without access to the hardware (McFarland 1971; McFarland & Houston 1981). The answer is that it is impossible to do this because of Kalman's (1963) controllability and observability theorems.

Briefly, in attempting to assess hypotheses, or models, we should ask whether the model contains redundant elements, and how many other models can also account for the data. If a model contains no redundant elements, it is said to be minimal. If a model is the only one that can account for the data, it is said to be unique. A model is minimal if and only if it is completely controllable and completely observable. A minimal model is unique. A model might be unique, but we could know this only if it was minimal. Behavioural models are never fully controllable or fully observable so they cannot be shown to be unique. For a particular behavioural model there is always a possible alternative that is just as good in accounting for the data (McFarland & Houston 1981).

The issue of passive versus active control is a live one in many fields. For example, many animals are able to regulate their body weight with considerable precision, and even growing animals will return to their normal (should be) weight after they have been forced to deviate from it. An active-control model has an internal representation (or set point) of the desired body weight which is compared with the actual weight; it provides a useful way of portraying the situation. It is difficult, however, to imagine that an animal has a device for telling itself how heavy it should be, or that it has a means of measuring its own weight. Considerable controversy has arisen over the use of the set-point concept in this context (Mrosovsky & Powley 1977; Reddinguis 1980; Wirtshafter & Davis, 1977), and one body of opinion holds that the concept has descriptive value only. The term "settling point" (Davis & Wirtshafter 1978) conveys the idea behind the alternative (passive-control) model, which involves physiological equilibria reached through complex interacting processes that are designed and calibrated (by natural selection) to balance each other (McFarland 1971; Toates 1980). Just as this body-weight control system "hangs together" in such a way that equilibria are restored, so any apparently goal-seeking system can be designed to hang together to achieve goals that are not represented in the system (McFarland & Houston 1981). [See also Toates: "Homeostasis and Drinking" *BBS* 2(1) 1979, and L. Magnen: "Dual Periodicity of Feeding in Rats" *BBS* 4(4) 1981.]

In considering apparently intentional behaviour we thus have a choice between models that postulate internal representations that are instrumental in guiding the behaviour, and models that claim that the system is so designed that the apparent goals are achieved by rule-following or self-optimising behaviour.

Optimising systems do not necessarily involve goal representations (McFarland & Houston 1981). Since we cannot tell (without inspecting the hardware) whether or not they do, we should recognise that both alternatives are viable (indeed they may be intertranslatable). The issue is not the simple one that Dennett implies, of cognitive versus behaviourist explanation. One issue is whether internal representations that control behaviour can be shown to exist. Another issue is whether such representations (if they exist) correspond to what we call intentions. The relationship between cognition and intention is yet another issue, and what all these have to do with communication is another issue again.

What then of the Sherlock Holmes method? This may be a convincing method for one who already believes in intentional systems, but a little caution is advisable. As a graduate student I used to experiment at making pigeons frustrated by presenting food on some occasions and not others, or by presenting food that could not be obtained by the bird. One particular pigeon used to behave rather aggressively toward me, and I formed the impression that he did not like my treatment of him. One day I inadvertently performed a Sherlock Holmes experiment and allowed the pigeon to escape from his cage. He immediately marched over to the tangle of electrical wires that controlled the apparatus and started to pull them apart with his beak. Feeling rather shaken and guilty I quit the room to obtain some coffee. Upon returning I realised, from his behaviour, that the pigeon regarded the wires as nest material, and that his aggression toward me was typical of the early stages of courtship.

Adaptation and satisficing

J. Maynard Smith

School of Biological Sciences, University of Sussex, Brighton BN1 9QG, Sussex, England

Like many scientists, I approve of philosophers if they reach conclusions that seem to support the way I do science, and to dislike them if they do not. By this criterion, Dennett is one of the best philosophers I have come across. His attitude to adaptation is almost exactly my own. However, he is more than that, because he has told me several things I did not know, and am glad to know.

I am only peripherally concerned with the debate between behaviourist and cognitive explanations of behaviour. Dennett's account of intentionality, in terms of the criterion of referential opacity and of an assumption of rationality, and his suggestion that specific hypotheses of intentionality can lead to testable predictions, seems convincing to an outsider. I'm sure it will not convince all behaviourists. However, I hope they will not reply that intentional hypotheses are not needed because all behaviour can be explained without them. After all, a similar argument would show that psychology itself, even behaviourism, is unnecessary, because all behaviour must ultimately be explicable in terms of physics and chemistry.

I have been more concerned with the problem of adaptation. I was delighted with the parallel Dennett draws between Skinner on the one hand and Gould and Lewontin on the other. I look forward to opening a seminar with the words "Harvard conservatives Gould and Lewontin." A few words on "satisficing": If I understand it rightly, the use of the term in economics is as follows. Suppose there is some policy A that maximises, say, output for some given set of inputs, and that a simpler rule of thumb, B, does almost but not quite as well. To adopt B would be to "satisfice." The justification for adopting B is that it is simpler, and less costly in management time and training. In other words, if one takes into account all inputs, including management costs, A is not optimal and B is.

The situation in animals is analogous. Tits, faced with the two-armed bandit problem, presumably do not use Pontryagin's maximum principle, but they do pretty well. One could, if one liked, say that they *could* solve the maximum problem, but that this would require too large and costly a brain. However, it seems more realistic to accept that there are limitations to what is possible to a given type of animal, and to seek an optimum subject to that constraint. Another reason for preferring the concept of optimisation subject to constraints to that of satisficing is that population geneticists are aware that even a very small difference in fitness between phenotypes will have evolutionary consequences.

I raise this point because there appears to be, among psychologists studying learning, a debate between supporters of the "matching law" and those who hold that animals learn the optimal behaviour. Surely this debate must, at least in part, be based on a misunderstanding? Suppose, for example, that in a learning experiment the red key was rewarded when the trial number was a prime, and the green when it was not. No one would expect a pigeon to learn the optimal behaviour; the best it could do would be to always peck the green key, because there are more nonprimes than primes. [See also Rachlin et al.: "Maximization Theory in Behavioural Psychology" *BBS* 4(3) 1981.]

This is a digression from the main thrust of Dennett's article. The important thing to remember is that, in using optimisation, we are not trying to confirm (or refute) the hypothesis that animals always optimise; we are trying to understand the selective forces that shaped their behaviour.

Parlez-vous baboon, Bwana Sherlock?

E. W. Menzel, Jr.
Department of Psychology, State University of New York at Stony Brook, Stony Brook, N.Y. 11794

Dennett's advice to students of animal communication and his concern for whether or not monkeys and apes really have societies, communication, intelligence, or intentionality, as humanists would define those things, leave me scratching my head with my foot. As Darwin put it, "Origin of man now proved. Metaphysics must flourish. He who understands baboon would do more for metaphysics than Locke" (as quoted by Sulloway 1979, p. 84).

A modernized translation of this sentiment would read about as follows. (I rely here most heavily on Mayr 1982.)

a. Philosophical and psychological concepts must be reformulated in the light of new evolutionary and behavioral evidence far more than the other way around. What we as humanists think humans do is not the yardstick by which all brands of mentality can best be gauged.

b. A necessary step toward understanding the workings of natural selection is to appreciate the uniqueness of individuals and the difference between populational or variational thinking and typological or essentialistic thinking. There are no "simple" animals. There is no fundamental essence of real communication, intentionality (and so forth) that living beings either "possess" or do not possess; that is confusing Platonic metaphors with reality.

c. The usefulness of "thought experiments" is limited to those problems in which all pertinent laws of nature may be assumed to be simple, perfectly reliable, eternal, and (at least in principle) transparent to the intellect. In the domain of ethology and comparative psychology, there are few if any such problems. (Things were of course different in earlier eras. Consider, for example, the Cartesian method, or Sherlock Holmes's smarter brother Mycroft, who outdid him without even getting out of bed.)

d. The aspiration for one-shot Sherlock Holmes-type litmus tests that will enable one to judge with certainty into which mental "class" or "type" any given set of individuals falls (without regard for variational problems) rests on the same or similar conceptions of nature and of scientific knowledge. This is not, of course, to say that there is not much room for improvement in scientific methodology. If one can reliably judge within seconds whether or not a dog sees one's approaching car or intends to cross the road before the car gets there, it does seem a bit absurd that if we were to take that same animal into the laboratory it might require weeks of work and 1,000 or more test trials before we could convince ourselves that it can discriminate black from white or shows anything at all analogous to "real intentionality." On the other hand, the difference between everyday problems and scientific problems must not be overlooked. Very different "courts" are involved; their "court rules" might be the opposite of one another in some cases (i.e. the "client" is assumed "guilty" until proved otherwise, or assumed "innocent" until proved otherwise, depending on the court); and the relative costs that one incurs if one's decision is wrong might not be measurable on the same general scale.

e. The most important and perhaps the only obligatory steps toward assessing an animal's so-called native intelligence are to ask how it manages to make its living in the world at large (especially in the face of novelty and change) and how it came to be capable of its feats. Insofar as these questions are ignored, the central problem of ethology and comparative psychology (cognitive or otherwise) has not even been addressed – regardless of how the animal performs on tests of our own devising. The difference between the study of natural intelligence and the study of artificial intelligence is that in the latter case these questions are not asked. Whether or not computers are really intelligent in the same sense as living beings is, by the same token, a non-Darwinian question. In Darwin's words, "Mind is function of body [and of genotype]. We must bring some stabile function to argue from" (as quoted by Sulloway 1979, p. 241).

As I read Dennett, he is either unclear on these issues or he (wrongly, in my opinion) confuses them with "behaviorism." Some brands of behaviorism are as typological and abiological as some brands of cognitivism; it pays to specify which brands one is talking about. (Konrad Lorenz is, coincidentally, just as "cognitive" as Tolman *when it comes to mammals;* indeed, apart from the fact that Lorenz is a naturalist and Tolman was a laboratory psychologist, the major difference between them is that at the empirical or data level of his research Lorenz is more interested in how the body parts of an individual move relative to one another, whereas Tolman was almost exclusively focused on where "the animal as a whole" was headed – i.e. the "purposive" or "directional" component of its activities.)

Dennett implies that his position is a novel one, and he argues that if future students of animal communication follow it they will see things that have thus far been overlooked. He may be right in the long run; but I did not detect anything in the target article for which one could not find clear precedents. Thus, for example, Köhler (1925; 1947) and Lorenz (1971) are still in my book the grand masters at practicing and expounding upon "one-shot" observations and demonstrations. Almost any textbook on learning theory or experimental psychology discusses how to test cognitive hypotheses or models; and what is any *single* trial in any experiment other than an "anecdote" of sorts (the typical lab experiment being a formalized way of collating many such anecdotes and planning their collection in advance)? Far from being different from the methods of other investigators, Dennett's method is an example of them (cf. Menzel 1969; 1979). Tolman (1951), Wiener (1946), Sommerhoff (1950), and Miller, Galanter, and Pribram (1960) not only show how teleological concepts can be dealt with scientifically, but do so without making bogus monsters out of those who choose to describe the same phenomena from alternative points of view. And I believe it fair to say that Menzel (1971; 1974) and Menzel and Johnson (1976; 1978), who applied the ideas of the foregoing investigators specifically to the problem of chimpanzee behavior, anticipated a very considerable portion of the ideas and phenomena that Dennett calls on future investigators to explore. Demonstrations of deception (without special "training") and the importance of context, and tests involving transparent "hiding places" are cases in point. I do not claim to have "discovered" these phenomena, let alone to have studied them exhaustively; that would be absurd.

The "stance" that I adopted in collecting the empirical data for these studies was, coincidentally, as uncommitted to either behavioristic or mentalistic scruples as I could get. My central problem was, quite simply, Where will each and every member of a group of six or eight chimpanzees go next in their one-acre enclosure; and why do they go there rather than elsewhere? (This is an expanded version of Tolman's "lone rat at the choice point of a maze" or a denatured and miniaturized version of Jane Goodall's "community of chimpanzees of the Gombe Stream Reserve" – take your pick.) The principal data consisted of a set of maps of the enclosure, on which the instantaneous positions of each and every animal were recorded at each 30-sec. interval. My job, as I saw it, was (i) to account for the maximum amount of the locational variance of these data with the minimum number of parameters and (ii) to interpret what this means, insofar as I could, using a vocabulary that would make sense to my colleagues and students – and to myself. To accomplish either of these tasks (especially the last one) I of course had to make some movies and "qualitative" notes as well as to collect the above map data – and to spend at least as much time in reading as in direct observation.

I do not claim that such an approach tells one any more about what chimpanzees are "really doing" than the maps of military generals tell one what warfare is really like. At the same time, it did enable me to see any number of phenomena that I had never seen before, and that I myself might never have surmised if I had tried to fathom the point of view of the "enlisted men" from the outset. It was, indeed, principally when I came to "explaining" the data (rather than in collecting it or even formulating and designing the experiments) that I felt "forced" to invoke cognitivistic and teleological terminology; and if my cognitivistic descriptions cannot in effect be translated back into a "pictorial" or "maplike" language from their verbal form (at least by those who have been fortunate enough to watch these remarkable animals for themselves), it is probably a sign that either I or the reader have lost contact (either with each other or with external reality).

I could be wrong, but in my opinion this is the way that most students of animal behavior work today, especially if they think of themselves first of all as chimpologists or baboonologists or beeologists or birdologists and secondarily if at all, as cognitivists or behaviorists or philosophers. The best observers of animals, in my opinion, are those who not only "define" their concepts with pictures or line drawings or some analogue thereof, rather than in words, but can think in this language; I aspire to this ability or at least to some degree of bilingualism, but somehow a description does not also *feel* right unless it is more than pictorial alone. Also, inasmuch as the way that chimpanzees (or any other species) move about is obviously different from the motions that we might expect if they were operating solely in accordance with the known principles of Brownian motion, gravity, or these plus those that apply to all living things, including plants, I do not see anything particularly wrong in hypothesizing that they have minds (or some more specific "type" of cognitive organization) and using such hypotheses as tools of "discovery" rather than post hoc interpretation. If this is what Dennett and cognitive ethologists are also trying to get across, I don't see what all the fuss is about.

Dennett's rational animals: And how behaviorism overlooked them

Ruth Garrett Millikan

Department of Philosophy, University of Connecticut, Storrs, Conn. 06268

Exactly *where* does Dennett's proposed research strategy, whereby we would look upon the more complex of nature's creatures with an eye to determining how "rational" they are, break with classical behaviorism?

Dennett's advice – to look upon these creatures as entertaining beliefs and desires and as drawing rational conclusions therefrom – should not be confused with the suggestion that we postulate, either ahead of time *or at all*, that these creatures have little representations in their heads by means of which they succeed in being (to some degree) rational (Dennett 1971; 1981c). The difference between Dennett and Skinner is not that the one wants to postulate certain little things inside the black box in order to account for its input–output dispositions whereas the other does not. The difference, I suggest, is in choosing what to consider "input" and what "output," and in the method of deciding *which* output was possibly determined by *which* input.

Dennett does not say here just what a "rational" animal is, beyond suggesting that it must be a creature that can be viewed as making *logical* practical inferences, given whatever beliefs and desires it has. But before looking elsewhere in Dennett's works, it is important to see that merely to view animals as if they were logic users would not in itself force us to move one step beyond behaviorism, or indeed, beyond a search for simple tropisms. For every simple tropism and every candidate for a behaviorist law *could* be modeled as an inference pattern. The bee that drags whatever has oleic acid on it out of the hive can be modeled as performing the inference "Whatever smells like oleic acid is dead; whatever is dead I should drag out of the hive; this smells like oleic acid; therefore I'll drag this out of the hive." True, the first premise of this inference is false when the nasty experimenter has put oleic acid on a live bee. So nature wired in a belief that was true only for the most part; would it have been worthwhile to do better? Similarly, any training that results in a conditioned response can be modeled as involving first an inductive inference (say, to "every time I push this bar, food appears") and then a deductive inference ("so, since I want food to appear now, I'll push this bar now"). By adding a thousand such tropisms to a thousand such conditioned responses, we would get an animal that could be modeled as continually making inferences, but with no enrichment of our most simplistic theories of animal behavior.

Being "rational," as Dennett uses this term here (he uses it more restrictedly in Dennett 1981b and 1981c), must include two features besides being modelable as a maker of multiple good or valid inferences.

1. A rational animal is one that (a) has the beliefs it "*ought to have*, given its perceptual capacities, its epistemic needs, and its biography" and (b) has the desires it "*ought to have*, given its biological needs and the most practicable means of satisfying them," where "ought to have" means "would have if it were ideally ensconced in its environmental niche" (taken from a slightly different context but to the same end, Dennett 1981b, pp. 42–43).

2. Moving toward rationality, Dennett holds, is moving toward *truly* having beliefs, which involves "giving individual belieflike states *more to do*, in effect, by providing more and different occasions for their derivation or deduction from other states, and by providing more and different occasions for them to serve as premises for further reasoning" (Dennett 1981c, p. 69).

Combine 1a with the thesis that there is "a perfectly real but very abstract commodity" called "semantic information" that is to be found at every location or location-over-time that an organism may inhabit, which takes numerous forms, and which concerns numerous things near *and sometimes very far*. Dennett is urging the ethologist to have an eye out for any such information of a kind normally available in the organism's environment that it might not be beyond the capacity of the organism to gather (given the gross anatomy of its perceptual organs) and that it might be useful for the organism to have given its needs. Any such information, without specifying the physical package in which it comes, is to be considered as, possibly, an input to the organism that can affect its output (e.g. information about the information that exists at another location, say, in a fellow monkey's head, *may* be available and being used by Sam-the-Vervet). Contrast behaviorism, which never attempted to specify the input to an organism in terms of total sensory stimulations (thank heaven!) but retreated to specification in terms of whatever myopically near events it occurred to the experimenter, in accordance with principles never reflected upon, to mention.

Combine 1b with the (reasonable) thesis that the behavioral output of an organism contains semantic "inverse" information (directions) about what will happen as a result of this activity. Any such information, if it is information about something it might be useful to the organism to have *done*, is to be considered as *possibly* a (purposeful) output of the organism (e.g. Sam *may* act in order that Tom should believe that Sam intends . . .). Contrast behaviorism, which early gave up the attempt to specify output as bodily movements, and retreated to specifying it in terms of whatever myopically near events caused by the organism it occurred to the experimenter to mention (e.g. bar pressing).

Point 2 adds that the fully rational animal would be able to use whatever bits of information it does use *in any combination* for the sake of fulfilling *any desires* that it might have. Its used information would not be tied to predetermined uses and unavailable for others. Contrast behaviorism, which preferred to see all cognitive change as learning to *do* something specific – as correlating input with output in a directly lawlike rather than a *rational* way.

The intentional stance faces backward

Howard Rachlin
Psychology Department, State University of New York, Stony Brook, N.Y. 11794

Mentalistic terms are used in many ways (for which reason their use in science is problematic). One way mental terms can be used is to convey uncertainty, as in: "I think it will rain today." For some purposes such a statement might be replaced with: "The probability that it will rain today is greater than 0.5." Obviously, the latter statement could not be substituted for the former for all purposes. For instance, "I think . . ., therefore some cognitive process is occurring inside me," or "I think . . ., therefore I am a human being and deserve to be treated as such," could not be replaced with probabilistic statements.

A syllogism depends, for the literal truth of its conclusions, on agreement with the premises. If one or more of the premises explicitly or implicitly denies such agreement then the syllogism cannot hold. So "All A's are B / All B's are C / All A's are C" is a good syllogism while 'All A's are B / John thinks all B's are C / All A's are C" is not a good syllogism (although its premises and conclusion may be true). The use of "John thinks" here implies that the reader need not assent to the second premise. One could substitute, "The probability that all A's are B is greater than 0.5" for "John thinks all A's are B," without altering the truth or falsity of any syllogism (about A and B) in which that premise appeared. Similar probabilistic statements could be substituted for the other statements involving mentalistic terms used by Dennett as examples of logical intentionality. Unlike the statements with mental terms, probabilistic statements are not commonly used to imply either a cognitive process occurring inside of or the moral elevation of an animal. The empirical truth or falsity of a statement such as "The probability that it will rain today is greater than 0.5" would be tested by establishing criteria for days, similar in critical respects to the present one, and whether it rained or not on those days. There is a good deal of room for argument about those criteria, but they are clearly nonmental. Dennett, by identifying intentionality with mentality, jumps from one use of mental terms – to convey uncertainty – to other uses – "How high can we human beings go?" – with no apparent justification.

Dennett claims that "the use of intentional idioms carries a presupposition or assumption of *rationality* in the creature or system to which the intentional states are attributed." But the statement "The parrot says, 'Polly wants a cracker'" is intentional by Dennett's criterion; it is "picked out by the logical test for referential opacity." But it does not intrinsically presuppose a mental state in the parrot. Again, Dennett confuses the intentionality of mental terms (a property they share with some nonmental terms) with other uses of those terms, one of which may be to imply rationality in animals or people.

Why does Dennett suppose that the intentional idiom presupposes rationality in the subject? One possible reason may be that the logic of the observer who attributes mental states to a subject is incomplete and needs to be completed by additional steps which are then inferred to occur in the mind of the subject. For instance, the bad syllogism, "All A's are B / John thinks all B's are C / All A's are C" could be made into a good (logically correct) syllogism with the addition of the premise, "Whatever John thinks is true." Perhaps this unstated premise is Dennett's reason for identifying intentionality with rationality on the part of John. It is not a good reason; the additional premise denies the uncertainty conveyed by the mental term in the first place. The operation would be like adding the premise, "All probabilities greater than 0.5 are equal to 1.0" to the analogous probability syllogism. Dennett has dragged in rationality on the tail of uncertainty and then thrown out uncertainty. It would have done just as well (and been much less confusing) to have avoided logical uncertainty in the first place by avoiding the use of mental terms.

Dennett's technique for determining the level of an intentional system is, as he admits, not well worked out. Eventually, it might establish strict behavioral criteria for the ascription of mental terms to animal behavior. If it did establish such criteria, it would be a form of molar (Tolmanian) behaviorism, familiar enough to psychologists. The success of such behaviorism would depend on how well those eventual cognitive terms ("cognitive talents") were worked into a system ("cognitive processes") that could serve to predict and control (i.e. explain) behavior. I suspect that, once a group of mental terms was sufficiently defined, sufficiently anchored in behavioral criteria ("weighted overshoes"), those terms would not suit Dennett for purposes of mental "promotion" and "demotion," and he would use others.

If Dennett sees the eventual use of mental terms without behavioral criteria, then psychologists will also be familiar with the method as that of the Chicago functionalist school (Small 1901): Establish a situation (a Sherlock Holmes situation), observe the animal's behavior in that situation, and then assign mental states to the animal according to the observer's own intuitive introspections ("these orders ascend what is *intuitively* a scale of intelligence").

The original maze (patterned after the Hampton Court maze) was used to study the "mental life of the rat" because it seemed to provide an ideal Sherlock Holmes situation. That technique failed because, without behavioral criteria, psychologists could not agree on what mental terms to use. One can imagine, in a Dennettian world, the arguments among ethologists and psychologists, each trying to "promote" the object of this study and "demote" all the others.

In his defense of the "Panglossian paradigm" Dennett identifies Gould and Lewontin, who criticize optimality theory on the grounds of its vagueness, with Skinner, who criticizes mentalism on the grounds of its vagueness. This is farfetched. Discomfort with vagueness is a common failing, especially among scientists.

It is possible to have an optimality theory that is not vague, that is behavioral (see Rachlin, Battalio, Kagel & Green 1981), and that is falsifiable (on the grounds of its usefulness in predicting and controlling behavior). Perhaps it is even possible to have a falsifiable mentalistic theory in which mental terms are not weighed down by behavioral criteria. But we have not had any so far. Dennett's "intentional stance" does not seem to rest any more firmly on the ground than those mentalistic stances that have previously faltered.

Intentionalist plovers or just dumb birds?

Carolyn A. Ristau
The Rockefeller University, New York, N.Y. 10021

Yes, Dennett is hedging about animal thinking by "taking stances" (see the commentaries in this issue by Churchland and by Griffin; see also Griffin 1981), carefully enclosing animal "mind" in quotation marks, and, in his closing remarks, implying "Do animals think?" is equivalent to asking "Is this the best of all possible worlds?"

In any case, I'm just going to take his ideas and use them for all they're worth. What are they worth? To be clear about it, I think they provide a fruitful approach to designing experiments and recognizing the import of certain field observations. But there are not insignificant problems encountered in the nitty-gritties of applying intentional analysis to field experiments.

Just now, we are studying that injury-feigning bird that is one object of Dennett's musings. As traditionally trained ethologists, we are concerned with the impact on injury feigning and other antipredator behaviors of variables such as stage of the nesting cycle. But we also attempt an intentional analysis of the bird's behavior (Ristau 1983). Let us frame a rendering of the first order of analysis as: The bird (a plover nesting on the beach) *wants* the predator (a fox) or other intruder (a human) to follow her or *wants* to lead the intruder away from the nest. We would therefore predict that the bird should *monitor* the intruder's activities to be certain of the direction the intruder is moving in and whether he is *paying attention*.

Furthermore, the bird should behave differently depending on whether or not the intruder is following her or paying attention to her. In essence, when the intruder stops following, we expect the bird to display more vigorously, or possibly to come closer to the intruder, to reattract the intruder's attention and then begin displaying again. We are led to ask how the bird might recognize that an intruder is paying attention, and thus we begin to gather data on such minute aspects of behavior as the intruder's direction of eye gaze. In short, we are led to ask questions and gather data whose significance may not have been apparent to us without an "intentional" stance.

Another example: Even shortly after a bird's eggs have been destroyed, she should cease displaying if she understands that she is, by her injury feigning, attempting to lure predators away from her eggs; if she is merely a hormonal reflex machine, she should continue to display. (This statement is not meant to deny the likely importance, even to an "intentionalist" plover, of hormonal influences on various aspects of behavior, such as the threshold of displaying.)

A first-order intentional analysis would also suggest that in the presence of an outside distraction on the beach that could attract the fox's attention, an intentionalist plover should no longer bother to perform her display. She should be reasoning that a fox attending to a bunch of bright balloons we've released or an aggressive interspecific interaction we've staged should *not* be attending to her chicks. What conclusions can we draw from the results of such an experiment? If the plover does *not* do any distraction display when the balloons appear, is she purposefully withholding her display because it is unnecessary, or was she so confused by the distracting event that she didn't display? If the plover does feign injury, that could be seriously demoting evidence that the plover doesn't know what she is doing and merely displays haphazardly in the presence of a fox. Or have we encountered a gap in intentionality, with the plover so overwrought with concern for her chicks that she displays to the fox anyway? Are we convinced that the plover saw the balloons; was she monitoring the fox's behavior so as to be capable of determining whether he was no longer attending to her chicks?

In this example, and indeed as a lingering problem, neither positive nor negative results from one instance help us to promote or demote an intentional analysis convincingly. We will need many diverse experiments and observations to lend support to either a romantic or a killjoy interpretation of a bird's injury feigning or any other being's possibly rational behavior. (We also seem to have slid into a second-order analysis when we have the plover monitoring the fox's attention – presumably a mind state – or are we still in the first order if we speak only in terms of the fox's *behavior*, namely the direction of his eye gaze?)

Can we conduct experiments about a second-order intentional analysis? Such an analysis might be framed as: The plover *wants* the fox to *believe* that:

1. she is injured
2. she would be good prey
3. her eggs or young are not located where they really are.

What predictions can we now make? If the predator has found the nest, the bird at least should not do "false brooding" as one of her displays, if she understands what false brooding appears to be to the mind of the fox. (False brooding designates a bird sitting at some location other than the nest, perhaps even fluffing the wings, as though to brood eggs in a ground nest. The display is quite convincing to a human, and supposedly knowledgeable people may well find themselves trundling after nonexistent nests.) With a predator at the nest, we might expect very vigorous distraction displays, if the plover is trying to attract the fox's attention, or, even more convincing, she might behave aggressively toward the predator. Other interpretations are possible. A predator at the nest is presumably a more fearsome stimulus than a predator away from the nest, and should cause more intense responses, such as broken-wing displays or physical attack, rather than less intense ones, such as false brooding.

The diversity of interpretations arises in part because intentional analysis is a stance, not a theory. Among many other implications, it is not complete and does not make specific predictions. We need auxiliary hypotheses.

Certain intriguing aspects of intentional analysis merit further development. The gaps in intentional systems experienced by adult humans as well as children and presumably nonhumans should be studied empirically to determine the conditions under which such gaps occur. They may provide useful insights into the evolution and ontogeny of cognition.

The leap from first- to second-order intentionality is large, as Dennett himself notes. It seems to raise the question at which order the notion of "self" becomes necessary. Is it possible to avoid a notion of self in the first order? Instead of a plover mentating "I want the fox to follow me" she could more amorphously be using "Want fox to follow this way." Yet, in the second order, once one posits beliefs to another creature, why not posit a sense of self, its own mind, as well? But how extraordinarily difficult to gather convincing evidence that the cognizing bird has a belief about another's belief rather than a belief about another's behavior. Almost every field experiment one can conjure up, from plovers recognizing which predators are dangerous at different stages of the nesting cycle, to responding differently when a predator is merely wandering by as compared to hungrily prowling for food, can be described in terms of plovers learning about behavioral sequences, not mind states, of a predator. Which to choose then, mind states or behaviors as interpretations? On what basis? Simplicity? Isn't an imputed mind state sometimes a simpler assumption for an organism to hold than a series of detailed behavioral predictions? "The fox *wants* my chicks, is searching *for* them" can be a convenient mental shorthand and may allow both experimenters and plovers to make predictions in novel circumstances or even bizarre experiments.

Yet I have a pervading sense that there is something wrong with intentional analysis, wrong in the "psychologically true" sense. Indeed Dennett is not suggesting this as a final answer in our grappling with the possible cognizing of animals. I simply cannot believe young children communicating with peers and adults, paying attention to each other, are framing or understanding third-order intentional statements. Neither do I intuit that animals are doing this. Am I? There must be some other shorthand that is actually working.

But until I know what it is, I gratefully take "intentional analysis" along to help me sort out observations and design experiments.

Intentions and adaptations

H. L. Roitblat
Department of Psychology, Columbia University, New York, N.Y. 10027

Dennett's description of intentionality is reminiscent of Tolman's (1932; 1959) approach to purposive behavior. They both argue that behavior is most fruitfully described at a level of integration reflecting its goal-directed nature. Concentration on lower levels of analysis, such as that of the muscle twitch, is less fruitful because it is less systematic, consistent, and informative than describing behavior in terms of organized goal-directed *Gestalten*.

Dennett proposes that we examine behavior from an intentional stance and suggests a number of criteria by which we should be able to judge whether an organism is behaving intentionally or merely as a reflexive stimulus-response automaton. The method involves a linguistic trick (referential opacity), that on the face of it is not of much use in the analysis of nonlinguistic animal behavior. The method also involves an analysis of the optionality of behavior.

In order to distinguish intentional causes of behavior from automatic causes, there must be a sufficient variety of available behaviors (sponges, for example, are unlikely to have intentions) and the behavior that is performed must be optional. Given a sufficient variety of potential behaviors, at least two kinds of optionality can be identified. Positive optionality is demonstrated when a behavior p that normally follows event q occurs in the absence of q. The vervet who gives the leopard call in the context of a heated battle demonstrates positive optionality because it gives the leopard call (p) without, presumably, having seen a leopard (q). Negative optionality is demonstrated when the same behavior fails to occur following event q. The lone vervet traveling between bands demonstrates negative optionality when it silently climbs a tree on spotting a leopard. Nonoptionality is demonstrated when p occurs only following q and whenever q occurs (p if and only if q). A behavior that is nonoptional cannot be intentional for the individual since it is controlled exclusively and exhaustively by q. Optional behaviors, in contrast, are controlled only partially by external situations, and partially by internal representations of various sorts. In practice, attempts to falsify or verify forms of optionality are often difficult. No finite observation can prove that a behavior is nonoptional. Similarly, as Dennett notes, by increasing the complexity of q, for example, by including special states, motivation, and the like in our description of the situation, it is possible to develop a description that is sufficient to account for the behavior nonoptionally, at least in an ad hoc manner. To embrace the intentional stance is not to reject scientific determinism.

Deception is important because, when the subject commits the deception, it indicates positive optionality. A behavior is optionally released from its usual goal and is instead used in achieving a new goal. It also indicates that q is not a necessary condition for the production of p. When the experimenter successfully deceives the subject, it means that the sufficient antecedent conditions have been identified. Hypotheses regarding alternative versions of q can then be compared.

While this method is useful for distinguishing zero-order intentionality (i.e. automatic response) from first-order (behavior controlled by its consequences), I cannot see how it could be usefully employed to distinguish higher-order levels from one another in nonlinguistic animals. How does one go about distinguishing whether Tom wants Sam to be in the trees, Tom wants Sam to believe he should be in the trees, or Tom wants Sam to believe that Tom wants Sam to run into the trees? We have no direct way to discover what is in Tom's mind except through behavior. Under ordinary circumstances, each of these intentions seems to result in exactly the same behavior, namely a leopard call, which is optionally connected to the presence of a leopard. Techniques based on the same logic as underlies referential opacity or the Sherlock Holmes method (i.e. an analysis of substitutability of conditions) may help in revealing more precisely how the subject represents the situation, but this methodology will require extensive development if it is to be practically useful. "Boy who cried wolf" productions, on the other hand, can only tell us that not only are leopard calls optionally produced, the hearer also responds optionally to them. Furthermore, whatever "credit" we want to give to Tom for intentionality, it is clear that leopard calls are adaptive in an evolutionary framework. Adaptation, whether in a proximate intentional or purposive manner, or in an ultimate evolutionary manner, is a sensible guiding principle when considered within a broad perspective of other concurrent influences on behavior.

Interesting behaviors within the intentional stance are those that are the product of an animal's representation of situations and particular events. Identification of the relevant features of those events and representations within such a framework will require the further development of techniques that are consistent with Dennett's proposed methodology as well as illuminative of the representations used by animals (see also Roitblat 1982).

ACKNOWLEDGMENT
Preparation of this commentary was supported by grant BNS82-03017 from the National Science Foundation and grant 1 RO1 MH37070-01 from the National Institute of Mental Health.

Content and consciousness versus the intentional stance

Alexander Rosenberg
Department of Philosophy, Syracuse University, Syracuse, N.Y. 13210

Dennett's *Content and Consciousness* (1969; henceforth *CC*) constitutes the single largest quantum jump in our understanding of the conceptual and methodological situation of the cognitive and behavioral sciences. It was there that the relations among "folk" psychology, centralist neuroscience, and behavioral theory were first fully and clearly mapped out. The conclusions of that analysis seem to be utterly at variance with the present defense of intentional systems in cognitive ethology. I shall limit myself to reporting the most pointed of these divergences, in the hope that Dennett can show why each of them represents either a misunderstanding of his original view, or an improvement upon it. I would not consider my strategy as any more significant than a piece of academic antiquarianism if I did not believe that in *CC* Dennett got it right.

The aim of Part I of *CC* is "to describe the relationship between the language of the mind and the language of the physical sciences" (p. 90). In its summary Dennett says the following:

> The centralist [the neuroscientist] is trying to relate certain Intentional explanations and descriptions with certain extensional explanations and descriptions, and the Intentional explanations that stand in need of this backing are nothing more than the rather imprecise opinions we express in ordinary language, in this case the opinion that Fido's desire for the steak is thwarted by his fear of the thin ice. If the centralist can say, roughly, that some feature of the dog's cerebral

activity accounts for his fear . . . of what he takes to be thin ice, he will be matching imprecision for imprecision, which is the best that can be hoped for.

Precision would be a desideratum if it allowed safe inferences to be drawn from particular ascriptions of content to subsequent ascriptions of content and eventual behavior, but in fact no such inferences at all can be drawn from a particular ascription. Since content is to be determined in part by the effects that are spawned by the event or state, the Intentional interpretation of the extensional description of an event or state cannot be used by itself as an engine of discovery to predict results not already discovered or predicted by the extensional theory. Ascriptions of content always presuppose specific predictions in the extensional account, and hence the Intentional level of explanation can itself have no predictive capacity. . . . The Ascription of content is thus always an *ex post facto* step, and the traffic between the extensional and intentional levels of explanation is all in one direction. (pp. 85–86)

The passage suggests that the intentional stance on its home base, the "folk" psychology of human behavior, is predictively sterile in respect of behavior, and that it has no payoff for the understanding of the nervous system either: The ascription of content is always ex post facto and can be made only after the behavior it eventuates in, and the traffic between the intentional and the neural approach is always only from the latter to the former. Yet in the target article Dennett argues that the aim of cognitive ethology, to provide an information-processing model of the nervous system, can be attained by employing this approach so stigmatized in *CC*. He tells us here that the intentional idiom of desire and belief is a suitably rigorous language for the description of cognitive competence, provided "we are careful about what we are doing and saying when we use ordinary terms like 'belief' and 'want'." But the argument of *CC* is that the intentional stance only submerges detail, and never avoids trivialization except through falsification; it never generates predictive power, and cannot suggest which brute questions to put to nature. *CC* allows us to give an intentional diagnosis of "an incident of great cleverness," but it denies that the diagnosis can provide any prognosis about the precise behavioral upshot of an intentional state. Yet, in the present paper Dennett says "the intentional stance is in effect an engine for generating or designing anecdotal circumstances – ruses, traps, and other intentionalistic litmus tests – and predicting their outcomes." But what is clear about the intentional stance at home, in the "folk" psychology of human behavior, is that its predictions are either false or highly indefinite and generic – never topographical – and, more important, never improvable as a class (even when improvable by Sherlock Holmesian techniques in particular cases). The most important achievement of *CC* was to explain why this is so, while breaking down the metaphysical barriers that this fact seemed to erect between human action and physical processes. Perhaps Dennett has an argument to show that in cognitive ethology, and in behavioral science generally, our aim should not be predictions of improving topographical accuracy – better and better descriptions of exactly what movements the subject will make, under specified conditions. But if he does have an argument for this claim, it will undercut his assertion that intentional "specs" set design constraints, for the intentional stance will not then provide the sort of constraints on the black box that are strong enough to choose between any of an indefinite number of alternative nonintentional or less intentional neural systems. The reasons for this are given in *CC*, and summarized in the passage quoted: the neuroscientist "will be matching imprecision for imprecision."

Dennett in fact recognizes the imprecision that results from adopting the intentional stance. He reasonably predicts that the stance will reveal vervets to exhibit mixed and confusing symptoms of intentionality, passing some tests at orders higher than those of others they fail. The reason given is that humans are like this as well. But these irregularities should tip us off that the intentional stance is a dead end for both men and monkeys, indeed for Martians and machines as well. It is certainly the case that as fascinated human beings, learning of the apparent similarity of vervets to ourselves in respect of "communication," we are bound to raise questions in the intentional idiom. But *CC* explains why these natural questions are also scientifically sterile. They reflect our desire to anthropomorphize other creatures. The trouble with transforming this desire into a research tactic is that in doing so we draw analogies from the unknown to the unknown. Our aim in studying vervets is in large measure to understand ourselves better. The naturalness of the intentional stance with respect to "simpler" creatures provides some reason to believe than in learning more about them we may be able to learn something of use in understanding human behavior and cognition, something we cannot learn about ourselves directly because we are more complex and because we cannot or will not treat ourselves, in laboratories or in the wild, in the way we are willing to treat vervets. We forgo the opportunity to learn more about them and about ourselves if our theory of vervet behavior is just the theory that has been employed since time immemorial to explain human behavior, a theory that has improved in no respects as long as we have been employing it. This makes all the more striking Dennett's argument for the intentional stance not as a theory that is true or well confirmed, but on instrumental grounds, as an approach with a practical payoff for prediction, and for the development of better theory. For whatever might be concluded about the realism, the truth, or the inevitability of the intentional stance with respect to human beings, its instrumental vices, its predictive weakness, and its theoretical fruitlessness seem undeniable.

CC offered an adaptationalist evolutionary theory to explain how purely physical phenomena can be upgraded to intentionality, while explaining why intentional descriptions have no specific consequences for physical phenomena. The reason Dennett was justified in helping himself to an adaptationalist approach is that there is an already known, independently confirmed theory that stands behind this and other instances of the adaptationalist stance. It is the theory of natural selection, a theory statable and improvable independent of adaptationalist commitments, which has been linked to other theories in biological science so as to provide increasingly accurate predictions and more and more detailed and unified explanations. The adaptationalist stance that it underwrites is of course susceptible to well-intentioned and *correctable* misapplication. But the tissue of misapplications on which Gould and Lewontin (1979) construct an attack on adaptationalism does not undermine the stance just because it is secured by an independently confirmed theory of natural selection. The disanalogy between the adaptationalist stance and the intentional one is that there is no theory to stand behind and secure the latter stance. This leaves it open to criticisms, like Skinner's, criticisms whose more cogent versions Dennett provides in *CC*, and whose force casts grave doubt on the relish here expressed for pouring old wine into new bottles, for explaining animal behavior as human action. [See forthcoming BBS special issue on the works of B. F. Skinner, *Ed.*]

Adaptationalism is a justifiable strategy across the range of biological phenomena because it is the product of a theory tested and confirmed in at least one important subdivision of the subject and with respect to at least one set of phenomena in its domain of objects. The same cannot be said of the intentional stance. What it requires is not the conviction by ethologists that they have a conceptual carte blanche to employ it; it requires the statement of an intentional theory (like one sketched by Bennett 1976, for instance) to provide it with the kinds of strengths Dennett says it has even without such a theory. Those who, with the Dennett of 1969, think no such theory is forthcoming, cannot in good conscience recommend the intentional stance to cognitive ethology. Instead we should follow P. M. Churchland (1981) and make them look the implications of eliminative materialism straight in the eye.

Steps toward an ethological science

Mark S. Seidenberg
Psychology Department, McGill University, Montreal, Que., Canada H3A 1B1

I doubt if Dennett's proposal to view lower animals as intentional systems will be controversial except to radical behaviorists (who might best be left to play in their own sandbox). What is important about his proposal is not merely the idea that animals can be considered as intentional systems, but also that this stance can provide the basis for an empirical, hypothesis-testing methodology. The rationality assumption provides a fruitful way to generate hypotheses concerning animal behavior only because these hypotheses (and other, nonintentional ones) can be tested in familiar ways against various kinds of behavioral evidence. Ethologists have been quite willing to let intentional vocabulary into their descriptions of animal behavior; in this regard, their efforts resemble 19th-century attempts to study "animal intelligence" (Boring 1950; Romanes 1882). In analyzing vervet monkey vocalizations, however, Dennett does exactly what ethologists typically don't do, namely generate a range of alternative interpretations which can be evaluated, at least in principle, with respect to relatively specific behavioral evidence.

The method that is typically observed in cognitive explanations of animal behavior (and perhaps in adaptationist explanations of evolution) relies upon a relatively simpler principle, namely the *consistency criterion:* Evidence consistent with a particular (cognitive) (adaptationist) explanation provides a sufficient basis for accepting it. This criterion sacrifices the hypothesis-testing component of Dennett's program – the idea that one can both *propose* a range of "intentional," "rational," or "adaptationist" accounts (as well as others) and *evaluate* them against relevant data and competing accounts. It is the latter component, however, that distinguishes these explanations from hand waving.

Much recent research on animal cognition accepts the intentionality assumption but fails to make good on the testing part. This is seen by considering a recent notorious, but not atypical, example: research on ape language. Apes such as Washoe or Koko are said to possess linguistic skills because their behavior is (weakly) consistent with this interpretation (Gardner & Gardner 1978; Patterson, 1978). Researchers offer high- (Washoe) or higher- (Koko) order interpretations of the behaviors in question, and they provide evidence (usually in the form of anecdotes) consistent with them. In these studies, alternative interpretations are not explicitly proposed or evaluated (Seidenberg & Petitto 1979; 1981); thus the evidence that could bear upon them is not solicited. This approach represents a radical departure from the usual hypothesis-testing methods of science (and from Dennett's proposal), but is wholly representative, in my opinion, of research in the ethological tradition.

I think this problem derives from the fact that while the consistency criterion is not a very good basis for doing science, it is nonetheless a prominent attribute of the naive psychology of explanation. Although use of the criterion is a familiar (and serviceable) characteristic of nontechnical explanations, the ethologists' unfortunate move was to import this method into "scientific" explanations of animal behavior (or perhaps introduce it to the 20th century). The similarity of the ethological approach and the naive psychology of everyday explanation is seen by comparing, for example, the methods of animal language researchers with those used in a somewhat more mundane context. William Safire writes a column for the *New York Times* that attempts to provide explanations for certain peculiarities in American English speech (e.g. the origin of the expression "fudging one's data"). Safire is a linguistic ethologist. His single methodological precept is the consistency criterion: An explanation is true if there is some anecdotal evidence consistent with it. Because the consistency criterion provides little constraint on what could qualify as a potential explanation, and no basis for deciding among plausible alternatives, Safire often faces the same problem as the ethologist or adaptationist, namely several competing explanations for a given phenomenon.

Thus, I would say that if cognitive ethology is to be taken as more than entertainment (à la Safire), it must incorporate not only Dennett's intentional stance, but also a methodology that evaluates more than a single hypothesis at a time.

Reservations: (1) Dennett's discussion of the role of anecdotes is not helpful. Anecdotes are isolated examples of behaviors which cannot be properly understood outside the context of a broader range of behavioral phenomena. The very most they can do is suggest that there might be an interesting phenomenon to be pursued in some more systematic fashion. When Dennett talks about creating conditions that provoke telltale behaviors, he is talking about doing experiments, not gathering anecdotes. Frequency of occurrence is not at issue; a behavior only has to occur often enough to allow anlaysis of the relevant conditions of occurrence, eliciting factors, consequences, and so on. A behavior so novel that it can't be observed more than once can't be understood.

2. Dennett has created a false dichotomy between his own views and those of Gould and Lewontin (1979). The "Panglossian paradigm" the latter described was the idea that for every behavior, there is necessarily an adequate adaptationist explanation. As they explain in some detail, theories of evolutionary biology offer numerous other possibilities, which must be played against any adaptationist scenario (but seldom are). Dennett wishes to assume the rationality of certain systems (people, computers, animals), and test hypotheses derived from this stance against nonintentional alternatives. The Panglossian version of Dennett's proposal would be quite different, namely: For every observed behavior, there is necessarily an adequate intentional explanation. But Dennett does not claim this at all, having rightly acknowledged that demoting evidence may be forthcoming in some cases. Thus, ethological explanations will not be exclusively intentional; evolutionary explanations will not be exclusively adaptationist. In each case, the strategy for formulating and evaluating alternatives is the same.

A better way to deal with selection

B. F. Skinner
Department of Psychology and Social Relations, Harvard University, Cambridge, Mass. 02138

Natural selection, operant conditioning, and the evolution of cultural practices are all examples of selection by consequences, a causal mode found only in living things. As I have pointed out in a paper that Dennett does not cite (Skinner 1981), they have features in common just because they exemplify selection but otherwise represent different biological and social processes. The *communication* that follows from natural selection is only superficially like the *communication* that results from operant conditioning. Vervet monkeys have evolved in such a way that when two or more of them are together, the one who first sees a predator emits a call in response to which the others take appropriate action. There is one call and one action for a leopard, another call and another action for a snake, and so on. The behavior of all parties has been genetically selected by its contribution to the survival of vervet monkeys. Speakers of English have been conditioned by a verbal community in such a way that when two or more of them are crossing a busy street, the one who sees a danger "emits a call" in response to which the others take appropriate action. There is one call for trucks, another for an open manhole. The behaviors of monkeys and people have similar features but are actually very different and must be analyzed in very different ways.

Dennett's intentional system theory and his use of words like *want, recognize, believe,* and *intend* appear to be an effort to recognize the role of consequences, but, as in the history of mentalism in general, they also allude to internal initiating causes – cognitive and intentional. We do not need them in accounting for innate behavior or for conditioned behavior in nonverbal organisms. Nor are they needed to explain behavior of young children or adults behaving unconsciously.

A hungry rat presses a lever and receives food. The probability of pressing is increased as shown by an increased rate of pressing. The rat cannot say "I press because I *intend* to get food, or because I *believe* I will get food, or because I *know* that I will get food this way," nor would it say so if it could, unless additional contingencies of reinforcement were arranged. A young child going to the cookie jar would only be puzzled if asked, "Are you going because you intend to get cookies, or because you believe or know there are cookies in the jar?" Additional verbal contingencies are needed before a child responds that way.

Dr. Pere Julià and I have been engaged for some time in an intensive analysis of several current philosophical concepts, including intention, belief, and knowledge. We feel that there is much to be gained by restricting the use of such terms to instances in which they can figure in first-person statements. They function, like the autoclitic – analyzed in my *Verbal behavior* (Skinner, 1957) – to promote more effective social behavior on the part of those who hear them.

The underlying terms in Dennett's statements 9–13 are apparently offered as referring to cognitive and intentional states or acts. They can all be interpreted as referring to consequences. For example:

9. x *wants* y to *believe* that x is hungry.

This would be true of a particular kind of "mand" (Skinner 1957) which is called a request because of the nature of the behavior of the listener. The customer says to the waitress "A ham sandwich, please." This is an order which the waitress fills because she is paid to do so. Statement 9, however, applies to a mand that is reinforced because the listener is disposed to give food to x as soon as x is discovered to be hungry. It is a request. Additional contingencies might lead x to say "I want a sandwich" and y to say "I believe you are hungry." But the verbal episode itself may occur in the absence of these self-descriptions.

The reinforcing consequences in operant behavior are usually observable and manipulable. In natural selection the role of survival can be demonstrated, but for the most part it is an inference, the inference upon which evolutionary theory rests. In both cases the consequences either have physical dimensions or may plausibly be assumed to have them. The dimensions of wanting, intending, recognizing, and so on, as initiating feelings or states of mind, have never been established, and the hope that neurology will eventually show that they are physical is no more than a hope.

Nonhuman intentional systems

H. S. Terrace

Department of Psychology, Columbia University, New York, N.Y. 10027

Understanding the causes of another organism's behavior is an exasperating problem even if we limit our concern to human behavior. Direct interrogation may provide helpful clues, but, as both Freud and Skinner have argued (though from obviously different premises), we are typically unaware of why we act, feel, and think as we do.

Though unhelpful in illuminating the causes of behavior, verbal descriptions of our thoughts and feelings can provide useful and relatively uncontroversial evidence as to their existence. It scarcely needs mention, however, that verbal interrogation can tell us little about the inner lives of nonhuman animals. Just the same, speculation about the psyches of various animals has preoccupied philosophers, biologists, and psychologists ever since Darwin provided a conceptual basis for raising such issues. It is not too much of an oversimplification to observe that the major goal of a century of research on animal behavior – by ethologists and behaviorists alike – was to provide a conceptual basis for defining an animal's knowledge of its world, and that the method of choice for discovering such knowledge was to evaluate an animal's responses to particular stimuli.

As Dennett notes, a major shortcoming of this tradition is its failure to consider *psychologically* interesting processes that occur between the stimulus and the response. In this respect, Dennett's point of view is similar to that of recent theoretical and empirical treatments of animal cognition (cf. Hulse, Fowler & Honig 1978; Roitblat 1982; Terrace 1982a; 1983). The "intentional stance" that Dennett proposes defines an interesting new approach to the study of the nonhuman mind, with respect to determining both whether such minds exist and, if they do, at what level of complexity.

Dennett argues that careful observation of animal communication can provide evidence concerning an animal's tendency to "believe that," "know that," "expect (that)," "want (it to be the case that)" and, more generally, to manifest other states. Though the present form of Dennett's proposal is specific to the intentional content of communicative acts, it should, in principle, be possible to apply it to the study of animal cognition in isolated animals.

As compared with more "romantic" approaches to the study of the animal mind, Dennett's goals are relatively conservative. Instead of simply *projecting* attributes of human thinking onto some aspect of an animal's behavior (cf. Patterson & Linden 1981; see Terrace 1982b for a critique of their position), or arguing that animal communication provides a "window to animal consciousness" or that the sheer variability of behavior is evidence of animal thinking (Griffin 1978; see Skinner 1931 for a critique of an earlier form of Griffin's position), Dennett proposes a framework for demoting or upgrading the complexity of an animal's communicative behavior so as to reveal the appropriate level of intentional complexity.

The essentials of Dennett's approach are revealed by his application of Grice's (1957; 1969) analysis of the intentional content of human communication to the intentional content of the alarm calls of the vervet monkey. With considerable ingenuity, Dennett reviews the empirical data obtained by Seyfarth, Cheney, and Marler (1980) and concludes that such alarm calls are of at least a first-order level of intentionality, that is, "Tom *wants* to cause Sam to run into the trees" by "some noise-making trick." Dennett also cites some anecdotal evidence that suggests that vervet monkeys are capable of a second-order level of intentionality: "Tom *wants* Sam to *believe* that there is a leopard [and that] he should run into the trees."

As romantic as it would be to accept either conclusion, some gaps in Dennett's application of Grice's intentional system leave me bogged down at the killjoy zero-order level of intentionality that Dennett seeks to transcend: "Tom (like other vervet monkeys) is prone to three flavors of anxiety or arousal: leopard anxiety, eagle anxiety, and snake anxiety. . . . The effects on others of these vocalizations have a happy trend, but it is all just tropism, in both utterer and audience."

How can we know whether Tom does, in fact, *want* Sam to know X? Since it has been reported that Tom makes an alarm call only when Sam is present, Dennett concludes that, in some sense, Tom actually wanted to address Sam. A more fundamental issue that Dennett does not address, however, is the *voluntary* status of the vervet monkey's alarm call. The observation that an isolated vervet monkey fails to make alarm calls does not speak to this question. Many species react silently in order to avoid detection by a predator. It is, therefore hardly that outlandish to assume the kind of "context" switches that Dennett dismisses as a behaviorist's knee-jerk interpretation of the observation that an isolated vervet monkey does not call out when it sees a leopard.

To show that an organism wants to do X, it is necessary to show that there are comparable circumstances in which it elects *not* to do X. Suppose, for example, that two men, John and Bill, were walking through a forest and that John saw that Bill was in danger of being attacked by a snake. If Bill were John's friend, or, to put it more generally, if John did not bear strongly hostile feelings toward Bill, one would expect John to say something like, "Look out Bill! A snake's about to attack you on your left side!" Under circumstances in which John wished harm to come to Bill, John might remain silent and leave Bill at the mercy of the snake.

Can a vervet monkey choose whether or not to make an alarm call? Would Tom, who notices that his rival Sam is in danger of being attacked by a snake, withhold the snake alarm call in order to eliminate Sam and thereby advance to a higher position in the dominance hierarchy of his group? Alternatively, one wonders whether Tom would refrain from making an alarm call, whose only function would be to alert vervet monkeys of a rival group of some imminent danger.[1]

At a more prosaic level, one might try some version of an "omission procedure" (see Sheffield 1965; Williams & Willimas 1969) to determine whether a vervet monkey could learn to refrain from making an alarm call under conditions in which that call is normally made. One might, for example, place the monkey in a situation in which an approaching leopard (as shown in a film) would be made to recede only in the absence of an alarm call. (Other monkeys would be present, but in locations that would prevent them from seeing the leopard.) Given the adaptive value of warning other members of the group of some imminent danger, it is unlikely that a vervet monkey could refrain from calling out, no matter how enticing its reward for doing so.

There are at least two grounds for questioning Dennett's view that a vervet monkey can assume the role of a "Gricean" audience: the function of the audience per se and the recognition of a particular audience. Just as a foraging bee will only perform its communicative dance if it has an audience (von Frisch 1967), it may be the case that a vervet monkey (for different adaptive reasons) will only make an alarm call if it has an audience. It does not follow, however, that either creatures's communicative act is anything more than the killjoy reflex that Dennett so zealously tries to dismiss.

Gricean communication not only assumes voluntary acts of communication, but an awareness of a *particular* audience. The fact that a vervet monkey can discriminate calls made by a particular member of its group does not imply that the sender of a signal is addressing a particular individual of that group. How do we know that Tom is addressing Sam, as opposed to some other member of the troop occupying the same area?

The absence of evidence that a vervet monkey's alarm calls are either voluntary or directed to particular individuals suggests that their communicative acts can be expressed by the following nonintentional rule: When danger$_i$ is seen and, when within shouting distance of other vervet monkeys, produce alarm call$_i$.

In this connection, it is interesting to ask how alarm calls of vervet monkeys differ from bird calls or from other familiar examples of animal communication. In each case there is a "vocabulary" of communicative acts. Birds have been observed to sing for the following reasons: (a) to attract a mate, (b) to signal the presence of a predator, (c) to ward off conspecifics about to intrude upon their territory, and (d) to call to their young. Bees communicate the location and quality of food sites near their hives by performing an appropriate variation of the "round" or "waggle" dances. Are bird songs and bee dances intentional acts?

At least in the case of bee communications, Dennett correctly rejects the concept of intentionality as superfluous. As desirable as it may be to "cast off the straightjacket of behaviorism" and seek a presentable . . . theoretical vocabulary," it seems counterproductive to borrow too freely from models of human intentionality when evaluating animal communication. Dennett should be lauded for devising a framework for studying a difficult and important problem. Having done so, he must await the outcome of the interesting experiments he suggests, such as the one that would employ hidden speakers to test the "disembodied" call hypothesis.

My guess is that these experiments will yield results that will require a substantial rewriting of the kinds of rules needed to define the different levels of intentionality suggested by Dennett. To expect otherwise would be to misinterpret the recent successes of laboratory studies which have, with considerable rigor, demonstrated the existence of cognitive processes in animals. However, as clearly as these studies have identified various instances of animal cognition, they provide no basis for concluding that animals think like human beings (see Terrace 1982a; 1983).

The most obvious reason for suspecting fundamental differences between human and nonhuman cognitive processes is the linguistic basis of human thought. Given the fundamental importance of language in human intentional systems, it seems foolhardy to ignore this fact when investigating the nature of intentional systems in nonhumans. Thus, even if it were possible to show that the alarm calls of the vervet monkey are voluntary and that a vervet monkey addresses its calls to particular members of its group, we would be left with the fascinating and nontrivial problem of defining the *nonlinguistic* representations of alarm calls in vervet monkey speakers and listeners.

ACKNOWLEDGMENT

The preparation of this commentary was supported by an NSF grant (BNS-82-02423).

NOTE

1. Dennett almost comes to grips with this issue in his analysis of the levels of intentionality that might apply to Seyfarth's anecdote concerning a vervet monkey who made a leopard alarm call while that monkey's group was losing ground to another group. The call in question seemed to be a false alarm. Unfortunately, Dennett left matters at the speculative but "dizzying heights of sophistication" as to the devious wiles that a vervet monkey is capable of manifesting – speculations that lack empirical support and that also beg the more basic question, Are vervet monkey alarm calls voluntary?

Author's Response

Taking the intentional stance seriously

Daniel C. Dennett

Department of Philosophy, Tufts University, Medford, Mass. 02155

Reading several dozen commentaries is certainly an intense learning experience. The attempt to compress my response to an appropriate length has left some thought-provoking points slighted, for which I hereby offer a blanket apology. I have tried to answer, at least by implication, every objection I disagreed with, so it can be assumed I agree with points I don't mention. I want to thank all the commentators for taking my proposals seriously, and I also want to thank them for a wealth of references that will keep me busy for a long time. In general, commentators chose to concentrate either on my proposals about intentional system theory in ethology and psychology, or on my Panglossian coda, and I have shaped my response to that division.

First, I respond to questions about the status of intentional system theory. Is it instrumentalistic, fictionalistic, old hat, not a theory at all – or just false? Second, I respond to skepticism about whether it is useful at all, or just a distraction. More specifically, what good, if any, is the Sherlock Holmes method? Third, I respond to particular objections and suggestions about the first part of the target article not covered in the first two sections. Fourth, I turn to the "Panglossian paradigm" and the debate about adaptationism.

I. Is it a theory, or what? Let us begin with "folk psychology," the "mentalistic" lore that is created by, and in turn helps shape, our practice of interpreting each other in daily life. Folk psychology has proven useful and efficient, at least for unscientific purposes, and, as **Humphrey** claims, it is eminently plausible to suppose that folk psychology is itself adaptive, an optimal (or close to optimal) method of behavioral interpretation – in the niche in which it evolved: prescientific and even prehistoric human culture. The evolution of folk psychology was probably an interaction of genetic and cultural evolution. Might the method of folk psychology be partly innate? It might (Stafford, unpublished). Just as the disposition to "consider as one's parent" the first large moving thing seen (to put it crudely) is genetically transmitted in geese, a disposition to respond to *any* large moving thing by "asking oneself" *what does it want?* would probably have survival value, since, for instance, distinguishing those moving things for which the answer is *it wants me* would be making a valuable discrimination. But even if folk psychology or mentalism has been useful up to now, we cannot project that it will continue to be the best method, now that we've added constraints and goals to the environment – not just the scientific goals of advancing psychology and neuroscience, but also the social goals of avoiding nuclear war, anticipating reactions to economic policy shifts, and so on. Are the good old-fashioned methods still best when we apply them in circumstances of such heightened complexity? That must be an open, empirical question.

But if this is so, we face great difficulty in exploring the question, since, as **Danto** points out the replacement in our actual practice of mentalistic folk psychology with an alternative is apparently unimaginable. We can imagine annihilating ourselves or turning ourselves into creatures incapable of sustaining a culture or science or society, but we cannot imagine continuing our lives as agents, discoverers, explorers, questioners, and scientists, without imagining ourselves continuing to be believers, desirers, expecters, and intenders.

Some – **Skinner**, Quine (1960), Paul Churchland (1981) – declare their independence from folk-psychological concepts, and in their various ways point to a future they cannot yet describe in which enlightened science will lead us to a new idiom. Here is one striking regard, then, in which Skinner (for one) is the very opposite of a conservative, for he (dimly) imagines a revolution overthrowing our outmoded ideology of mentalism, our heritage of Cartesianism, our false consciousness consciousness, one might say. The rest of us, in our various and often competing ways, are convinced that the intentional idioms are here to stay, at least for human beings, and try to accommodate them one way or another to the advancing edge of biology.

My view is that no simple, direct, "reductionistic" accommodation can be made – a view I share with many – and that the best sense can be made of folk psychology (of belief and desire talk in particular) if it is viewed *instrumentally*. So I am an "instrumentalist" – but not a *fictionalist* as **Churchland** and **Graham** would have it. Attributions of belief and desire are not just "convenient fictions"; there are plenty of honest-to-goodness instrumentalist *truths*. (Graham's commentary in particular is vitiated by this misinterpretation, but my own early imprecision no doubt deserves most of the blame for this misreading, which I have recently gone out of my way to disavow (Dennett 1981c). Consider the truths one can assert regarding an instrumentalist entity such as a *center of gravity:*

> As you slide the lamp out over the edge of the table, it will remain upright so long as its center of gravity is located over a point on its base still on the table.
>
> You can move the center of gravity of the lamp down by filling its base with water, and to the side by sticking a wad of chewing gum on the side.

Are centers of gravity fictions? In one sense, perhaps, but there are plenty of true, valuable, empirically testable things one can say with the help of the term – and one doesn't fret about not being able to "reduce" an object's center of gravity to some particle or other physical part of the lamp. *Explanations* may refer to centers of gravity. Why didn't the doll tip over? Because its center of gravity was so low. (This explanation is not obviously "causal" but it surely competes with others that are: "Because it is glued to the table." "Because it is suspended by invisible threads.")

I want to claim much the same sort of thing about belief claims. You can change the monkey's belief from p to not-p by doing such and such; so long as it believes that p and desires that q, it won't try to do A, and so on. Why did the monkey look in box B? Because it believed there was a banana in box B. This intentional explanation competes with other explanations, such as "Because it had been conditioned to look in box B whenever a bell rang, and a bell just rang." Just as there are physical facts in virtue of which a lamp's center of gravity is where it is, so there are physical facts in virtue of which a monkey believes what it believes. But let us not be too impatient to declare exactly what shape those physical facts will take in general. I decline to identify beliefs with any "causally active inner states of the organism" (**Churchland**) for the same reason I decline to identify the lamp's center of gravity with any such inner state or particle.

Is this instrumentalism immune to falsification, as **Churchland** claims? Particular attributions of belief and desire are certainly falsifiable, as I showed in the target article (pace **Graham**). But it is indeed very hard to imagine what could overthrow or refute the whole scheme of belief and desire attribution, for by its instrumentalism it avoids premature commitment to any particular mechanistic implementation. This is a strength, not a weakness.

Lloyd also overstates my instrumentalism. While I do indeed think that what we human beings share with thermostats (and yes, even shortest-path-seeking lightning bolts) is worth elevating to attention, I also insist on the differences. Rationales are *not* "*always* free floating." The more complex, interesting, versatile an intentional system is, the more inescapable it becomes to interpret its innards as involving systems of representation. (**Millikan** quotes from the relevant passage in Dennett 1981b.) It is precisely for the *indirect* light that intentional system characterizations shed on these systems of representation that they are so useful to science, and not just as guides for social interaction.

The need for this indirection, and the complexity of the issues deliberately submerged by intentional system theory, is brought out by **McFarland.** Optimizing systems are systems predictable from the intentional stance, but, as he points out, and illustrates with the example of body-weight maintenance, optimizing systems do not necessarily involve goal representations. "In considering apparently intentional behaviour we thus have a choice between models that postulate internal representations that are instrumental in guiding the behaviour, and models that claim that the system is so designed that the apparent goals are achieved by rule-following or self-optimising behaviour." Now McFarland assumes, plausibly but mistakenly, that I intend to restrict the class of intentional systems to only the former sort, the "active-control" systems, in his terms, those systems that contain "a representation of the goal (or want)." *Eventually,* I grant, we need a theory that breaks people and other organisms down into what I gather McFarland would call their active-control and passive-control subsystems and subprocesses. That is, we want to work toward an account of internal processes that will distinguish between those cases in which a particular representation plays a role and those in which the information is only virtually or tacitly present in the design of the system. (See Stabler 1983 for an acute discussion of this issue in linguistics, and Dennett 1983 for further groping in this direction.) As **Harman** correctly notes in his point 4 (see also **Bennett**), we must also distinguish between what an organism knows or believes and what it *relies on* in the instance. (See also Harman 1973 and, for a strikingly different perspective, Millikan, forthcoming.)

So while I do not at all deny that we should strive for a theory of actual internal information processing, a theory of the "causally active inner states" **Churchland** mentions, and while I would also insist that the first elements of that theory are beginning to emerge from cognitivist research, my point is that one should not confuse *the predictive success of the intentional stance* (in some domain) with *confirmation of a particular representation-manipulation hypothesis.*

Most of the references cited by **McFarland** are new to me. They seem to address issues of central puzzlement to me, and I look forward to reading them, but have been unable to do this in the time limits set by this BBS Commentary. Particularly striking is his claim that McFarland and Houston (1981) show that "any apparently goal-seeking system can be designed to hang together to achieve goals that are not represented in the system." *Any?* Any number of different goals in one system? Goals with indefinitely sophisticated satisfaction conditions? It sounds like a proof that the Rylean dream of the completely representation-free realization of an intentional system is possible. That is too much for me to swallow at this point, but it will certainly be interesting to see what neighboring hypotheses are defended.

It is often illuminating to move issues back and forth between their intentionalist and adaptationist arenas – one of the themes of the target article – and here is a case in point. My reasons for recommending that we understand intentional system theory instrumentally can be clarified by considering what the counterpart would be in evolutionary theory. Imagine a world in which *actual* hands supplemented the "hidden hand" of natural selection, a world in which natural selection had been aided and abetted over the eons by tinkering, far-sighted, reason-representing, organism designers, like the animal and plant breeders of our actual world, but not restricting themselves to "domesticated" organisms designed for human use. These bioengineers would have actually formulated, and represented, and acted on, the rationales of their designs – just like automobile designers. Now would their handiwork be detectable by biologists in that world? Would their products be distinguishable from the products of an agentless, unrepresenting purely Darwinian winnowing where all the rationales were free floating? They might be, of course (e.g. if some organisms came with service manuals attached), but they might not be, if the engineers chose to conceal their interventions as best they could. (**Lloyd**'s reflections on those occasions when Mother Nature proves too smart for the adaptationists suggest that truly *bad* design that looked good at first to design critics might be the best telltale clue of human intervention – but of course that particular sort of clue would normally have a short half-life.) This is my point: A great deal of sound, productive adaptationist research on a species, its evolution, and its relation to its environment could be accomplished prior to, and independent of, any settling of the question of whether the species had *representations* of the reasons for its design in its ancestry. Eventually we would hope our theories could uncover the historical truth about these etiological details, but our hope might often be forlorn; there might be insufficient trace left for *any* science to be able to interpret. (If biology had to restrict itself to answering such etiological questions about the past, it might simply not be possible; it sometimes seems to me as if this canon and its nihilistic implication are embraced by **Lewontin.**)

This delicate relationship between *causes* and *reasons* is at the heart of **Rosenberg**'s commentary, which laments my backsliding from what he takes to be my major insight into the relationship between folk psychology and biology in *Content and Consciousness* (1969). Certainly the contrast he draws between my view then (to which he gives a fair interpretation) and my view in the target article is striking. What gives?

What I dimly saw in 1969 was what today I would call the *impotence of content,* but I misdescribed it slightly then. If meaning were an independent force or property or feature of things such that it could itself play a causal role, then a certain sort of predictive strategy should be possible: Determine *exactly* what the meaning or content of some state or event was (exactly what A believed and desired, exactly what the message really means), and then calculate from this its effect on the rest of the world. But meaning is not such a causal property. There couldn't be direct meaning transducers, for instance. So *that mode* of predictive strategy is an illusory goal. But I overstated the case: I said that intentional (meaning-attributing) characterizations were, as **Rosenberg** puts it, "predictively sterile." They are not, obviously; nothing is more facilely and prodigiously predictive than the intentional stance. But intentional stance predictions are peculiarly *vulnerable;* they have no predictive *hegemony* over design stance or physical stance predictions – precisely because the meaning or content they attribute is not an independent causal property of anything, but a dependent, supervenient, approximated property. Even if we could always say what someone who believed exactly *this* and desired exactly *that* would do (*ought* to do), only that person's subsequent performance (or performance dispositions calculated at the design or physical level) would show how close to believing and desiring exactly *this* and *that* the person was (Cf. Bennett 1976, sec. 36, "What exactly does he think?")

So the radical view **Rosenberg** admired in *Content and Consciousness* became the more tempered view of "Intentional Systems" (1971), in which the intentional stance *is* viewed as an "engine of discovery," because it does give the "specs" of information sensitivity of the organism's biologically embodied control system. Rosenberg notes, correctly, that adopting the stance does not move one directly in the direction of providing "better and better descriptions of exactly what movements the subject will make, under specified conditions." That is too hard a task for now. The intentional stance makes life easier for the scientist by characterizing broad equivalence classes of *action* types to predict (Dennett 1978a, chap. 15; 1981c), and does, as Rosenberg claims, leave many importantly different accounts of internal operation indistinguishable – **McFarland**'s point. In fact I stress that fact in a deliberately provocative way in "Intentional systems" by pointing out that there is a sense in which intentional systems theory is "vacuous as psychology," precisely because it presupposes rationality. Similarly, the intentional stance explanation of a particular chess computer's moves ("it castled because it anticipated the discovered check if . . . ") is manifestly vacuous as computer science. But it is exactly the way to organize one's task before doing the nonvacuous, nontrivial design work.

Moving from a description of competence to a performance model requires increasing specification. By the time the topic turns to search trees, data structures, and evaluation algorithms, there is all the precision and rigor one could ask for. In between this subpersonal account of processes and the loose-fitting intentional systems account in terms of beliefs and desires, there is room for intermediate levels of modeling – for instance, flow charts and systems of rules to be followed by (but not necessarily represented in) the organism. (See, e.g., Newell 1982.) Is this the level of precise "theory" **Bennett** urges ethologists to aspire to? If it is, I would heartily concur, but I am not sure this is what Bennett has in mind.

Bennett claims that I fail to provide the "firm underlying theory" about "conceptual structures" required by cognitive ethology. Just such a theory has been attempted in Bennett (1976), a sketch of which is given in his commentary. Bennett's book is indeed full of insights that should be of interest and value to ethologists; in fact it discusses, in greater detail, virtually every topic of the target article. (Embarrassing note: Bennett and I, working entirely independently, arrived at a slew of similar conclusions at about the same time; it took our students and colleagues to put us in touch with each other's work a year or so ago. Now if there turns out to be someone named Cennett!)

Bennett grants that my "conclusions" are acceptable to him. Moreover, he is not claiming (so far as I can see) that his theory permits explanations, predictions, or verdicts that are inaccessible to me, given my way of doing business. Indeed, the accounts he provides in his commentary (e.g. of when and why to talk of the goal of leopard avoidance, what settles the issue of whether a high-order attribution to Tom is correct) are very much what I would have said, and to some extent have said on other occasions. The difference is that he claims to derive his conclusions the hard (and proper) way – from a rigorous, precise, articulated theory of conceptual structures – while I obtain the same results by what seems in contrast to be a slapdash, informal sort of thinking that I explicitly deny to be a theory in the strict sense of the term. Bertrand Russell (1919) once excoriated a rival account by noting it had all the advantages of theft over honest toil; Bennett, I am grateful to say, finds a variation on this theme: I stand accused of poetry.

I plead *nolo contendere,* for it seems to me that, aside from differences in expository style and organization, **Bennett** and I are not just arriving at the same conclusions (for the most part); we are *doing the same thing.* If Bennett has a theory, it is not – had better not be, for the reasons just reviewed – a theory directly about internal processes. The sort of behavioral evidence he relies on to anchor his claims simply won't carry theory that far. So his theory is, like my instrumentalism, a theory of "conceptual structures," as he says. The methodological difference I see is strictly in the format of presentation, with Bennett's theory being, like many other philosophical theories, "a system of definitions propounded and defended" (Shwayder 1965). I think the idea that there is a proper theory to be developed here is a philosophical fantasy. Getting clear about something does not *always* mean producing a clear theory of it – unless we mean something quite strange by "theory." (I stand in awe of the systematic knowledge about automobiles good mechanics and automotive engineers have, but I don't think they have or need a theory of automobiles – certainly not a theory that yields formal definitions of the main concepts of their trade.)

Let us consider one of **Bennett**'s examples of theory. We agree that the applicability of the terms "belief" and "desire" will have some "tailing off" or attenuation as we move down the complexity scale from *Homo sapiens;* I am content to speak of (attenuated) animal beliefs and desires; Bennett introduces a technical term, "registration," of which beliefs proper are an exalted species. The main *differentia* of beliefs are that in order to believe, and not merely register, that such and such, one must be "highly educable" and "inquisitive" about many similar matters (Bennett 1976). This is to distinguish the hard-wired or obsessive or single-track information-retention of lesser species from our more versatile sort. These are, surely, the most important differences between my way of registering that there is nectar at location L, and some honeybee's way of registering (roughly) the same fact; and it is just these differences the ethologist should attend to (see Gould & Gould 1982). But the formal rigor of a definition of "*a* believes that *p*" in terms of a previously defined concept of registration cannot usefully survive the inclusion in the definiens of such phrases as "highly educable" and "inquisitive." Everywhere one turns one finds matters of degree. As Bennett himself observes, "belief shades off smoothly into mere registration" (1976, p. 88). So having paid a heavy price in "poetry" for rigorous expression, we then discover that our every application of the technical terms is hedged with matters of judgment, *ceteris paribus* clauses (cf. Lewontin 1978a on the role of *ceteris paribus* clauses in adaptationism), and degrees of this and that. To me, these are the telltale signs of philosophical makework, a definitional tour de force that never actually gets used for anything – even by Bennett.

Note, too, that no sooner does **Bennett** introduce some of his technical terms than he excuses himself for committing a little bit of poetry, and lapses back into the vernacular – so he can actually make a point someone might follow. (Having said that, I must also remind nonphilosophers that, *as in their own fields*, a lot of the best work in philosophy is not readily accessible to outsiders, and often consists of projects of only intermittent interest to workers in other fields. Some philosophers have recently overcome their traditional condescension, done their homework, and learned a lot from other fields; people in other fields can find similar benefits in philosophy.)

I think ethologists should read **Bennett,** and then ask what benefits accrue if they take their medicine and do things his way. The proof, of course, will be in actual practice, and here Bennett does present one point of clear disagreement with me. I have advertised the "Sherlock Holmes" method of contrived anecdote provocation, but Bennett thinks my description of the method misleading and the method itself no advance. His alternative is apparently good old-fashioned "nomological-deductive" hypothesis testing.

> Despite what Dennett says, this is not a move from regularities to anecdotes; rather, it is a move from regularities of one kind to regularities of another. If the work is done right, there may indeed be "control," but that is not what makes the procedure "scientifically admissible." There is *no reason in principle* [my italics] why we should not make the enlarged set of observations with our hands behind our backs, not contriving anything but just looking in the right direction. The procedure is scientifically admissible just because it consists in objectively attending to data in the light of a decent hypothesis.

No reason in principle, but how, except in philosophers' imaginations, are we to *gather data* about the "regularities" of this and other kinds? Whereas very simple creatures can be treated by scientists more or less as if they were ahistorical specimens of this or that type, people cannot be forced time and again into these situations, and neither can monkeys

II. Does it work? The problem is analogous to the problem facing the historian: How could one test a hypothesis about the causes of the Crimean War? One cannot replicate, in one's world-lab, the relevant control experiments, for they involve "subjects" so complex, and so massively and intricately a product of their histories, that they can never be put in the "same state" twice, let alone many times, with many controlled variations. Now if **Seidenberg** were right, history would be an utterly forlorn enterprise, for he says, in expressing his reservations about the Sherlock Holmes method: "A behavior so novel that it can't be observed more than once can't be understood." My point is that it can, if one makes use of the intentional stance. The best one can do is to devise a *narrative interpretation* of the phenomena, and if it is a good one it will be able to yield predictions of otherwise "unexpected" turns of events.

The point of the Sherlock Holmes method is to predescribe circumstances and an effect in those circumstances that is predictable only by a certain intentional characterization. If the prediction is borne out, this is not absolutely certain confirmation of the hypothesis (a red herring raised by **Menzel**). One gets confirmation, in the end, only by varying circumstances; only by seeing what happens in a variety of cases. (See **McFarland**'s quotation of Lloyd Morgan, which makes perfectly the complementary point about the inconclusiveness of one-shot demotions.) Because of the way complex, learning organisms reflect their histories, however, this variety cannot as a matter of practical necessity be achieved by classic controlled variations on the first case. We can't test a child's comprehension of a story by reading it to him a hundred times with minor, controlled variations, and unless vervet monkeys are stupider than we think, we cannot expect good results from trying to get vervet Tom to perform his apparently clever deception on the rival band a hundred times in a row as we vary the circumstances. But still, we must do something to assure ourselves that the *apparently* clever act wasn't a dumb luck coincidence. *One* apparently clever act may well be a coincidence, but if we can often or regularly evoke wise or tricky acts in different circumstances, we will be ready to concede real cleverness. So in the Sherlock Holmes method one tries to steer the narrative – to get a particular sort of history to happen freshly, and include a "response" or action that has only one plausible interpretation. While **Menzel** and **Ghiselin** are certainly right, then, that an anecdote is just another observation, and while in one sense the Sherlock Holmes method is just a special case of classic experimental design (**Seidenberg**), it is a rather special case.

No doubt **Dawkins** is right that I have a long way to go before I convince ethologists that this is a trick worth trying; I would expect, for the reasons just mentioned, that the "higher" and and more complex a species, the more useful leverage the method would provide. **Heil** presents the case of a pit-digging trapper and wonders what it would take to establish that the trapper had beliefs about the beliefs of its prey, and not just beliefs about its likely behavior. If the trapper in question is an insect, one can certainly run the sorts of repetition-with-variation experiments that are not really examples of the Sherlock Holmes method. But if the trapper is a human being, other sorts of data are available – and required. Consider for instance, what can be fairly conclusively deduced about the trapper's beliefs and desires when one comes across a very carefully concealed steel trap in the woods under a large sign that says "DANGER. BEAR TRAP. KEEP AWAY."

I have no idea how disappointing in practice the method might be to ethologists. It is certainly subject to pitfalls – of the sort pointed out by **McFarland** – and difficulties – of the sort pointed out by **Ristau.** But as Ristau's current research reveals, it does generate lots of questions one can begin thinking about how to answer. For instance, Ristau asks about her distraction-displaying plovers: Does the plover monitor the fox's attention, or just its behavior (the direction of its gaze)? This is another opportunity to relegate a rationale to the free-floating category. If we discover by suitable tests that the plover relies stupidly on eye-gaze direction, we will not credit *it* with appreciating the rationale that ties eye gaze to the predator's attention, but the rationale is still a good one, and it would be passed to the evolutionist for explanation (see Dennett 1980). Ristau is currently experimenting with a radio-controlled stuffed-raccoon surrogate predator on wheels that can change its movement direction. Will ingenious tests of this sort be fruitful? I couldn't ask for a better trial. If it demonstrates in due course that this attention-focusing power of the intentional stance does more harm than good, then I will certainly have been shown to be wrong – one more case of a philosopher leading a scientist down the garden path. Caveat emptor. But if **Seidenberg** is right in his charge that ethologists have tended to settle for the all too weak *consistency criterion,* then the method offers some relatively novel leverage for *dis*confirmation of hypotheses.

III. Other points. The interpretation of animal messages can, **Griffin** says, tell us at least part of what the animals are thinking and feeling. I agree. The first step must be "radical translation" of these alien modes of communication, and for that task I know of no other method than the intentional stance. But I don't hold out as much hope for the fruits of this sympathetic listening as Griffin does. I think we already know enough about the environmental predicaments and corresponding talents of lower creatures to know that they have no use for the sorts of communicative genres that would have to exist before there could be a Proust-porpoise or Joyce-bat with much to tell us – or each other – what it was like to be them. Beatrix Potter's animals have a lot to say about their lives, but their lives are human lives. While I disagree with Wittgenstein's oft-quoted pronouncement – "If a lion could talk, we could not understand him" (Wittgenstein 1958, p. 223e) – I do think we'd find the lion had much less to say about its life than we could already say about it from our own observation. Compare the question: What is it like to be a human infant? My killjoy answer would be that it isn't like very much. How do I know? I don't "know," of course, but my even more killjoy answer is that on my view of consciousness, it arises when there is work for it to do, and the preeminent work of consciousness is dependent on sophisticated language-using activities.

Premack's point, quoted by **Jolly,** that chimps aren't smart enough to be behaviorists is excellent, I think, and underlines **Humphrey**'s claim that for simplifying, unifying power, it is hard to imagine what could beat mentalism as a way of understanding (or at least *seeming* to understand) the things that move around us. **Ristau** also makes this suggestion, and Jolly observes correctly that the author of TALESPIN was relying on the unparalleled organizing power of the intentional stance in treating gravity as an agent. This is an example of the ubiquitous AI (artificial intelligence) practice of organizing functional decompositions around homunculi or "demons" – minimal intentional systems that can be assumed to perform certain roles.

But I also agree with **Heil, Jolly,** and **Ghiselin** that to answer the question of just which hypothesis is parsimonious, and why, solely in terms of order of intentional iteration would be unsatisfactory. That is just one measure to be played off against others. (Parsimony is also a trickier matter than Ghiselin seems to think; his "*logically* simpler explanation," as a rendering of parsimony, is simply misnamed – unless he can give a *logical* definition of when a subsidiary hypothesis is ad hoc. If he can do that, philosophers will certainly pay attention.)

The role of the rationality assumption is questioned by some, but is nicely revealed in **Heil**'s discussion of someone skating on a frozen pond. Heil gives several rival candidate interpretations, and it is clear that indefinitely many others could easily be concocted. Note what holds each of them together, however, and in fact plays a major role in generating them: a coherence constraint. We don't attribute to the skater the belief that the ice is quite thin *unless* we attribute to him either a desire to get wet or to drown, or any one of the infinity of beliefs that would have the implication that in spite of the thinness, he won't break through: His faith will keep him up; he can fly; it is so cold that open water would freeze instantaneously, and so on. The parsimonious interpretation is, as Heil notes, the one that provides the most coherent rationalization of the skater's experience and behavior (see **Millikan**). But this very fact undercuts the initially plausible and straightforward line Heil takes about rationality: "S may hold p and hold as well (and with good reason) not $-q$. S may, for example, not *recognize* that p implies q." But not recognizing this is a way of falling short of believing exactly p. We do fall short in just this way all the time; hence the ideality of intentional system theory, and the riskiness of its predictive power (see the reply above to **Rosenberg**). (On the difficulties attendant on empirical tests of the rationality assumption, see Cohen 1981 and Kyburg 1983.)

Menzel issues obiter dicta (a-e); they are offered with no supporting argument, but insofar as I see their relevance, I agree with what he seems to be saying in them, and do not see that I have been unclear about the issues raised. He doubts that there is anything new in my proposals; Kohler, Tolman, Lorenz, Wiener, and others have taken care of all these matters quite well. While I have learned a lot from all of these authors, it is my impression that they got a few things wrong, and missed a few tricks. For instance, Tolman, whom **Roitblat** and

Rachlin also correctly cite as a forerunner, got bogged down on a positivistic and atomistic criterion hunt: trying to provide *piecemeal* "operational definitions" of intentional terms such as "the rat expects food at location L." This is just what couldn't work, as Taylor (1964, esp. pp. 76–82) showed with admirable care, scholarship, and insight.

As for **Menzel**'s own work, I must admit I had not come across it yet in my initial forays into behavioral biology, though from his cursory description it sounds interesting indeed, directly relevant to the topics I have been working on, and something I intend to study.

Skinner sees no advantage to be gained by adopting a mentalistic idiom, but his own claims hover equivocally between demonstrably false behavioristic interpretations. I would like to concentrate on one highly characteristic claim of Skinner's which exhibits the familiar problems with his brand of behaviorism.

Skinner is confident he knows the true account of the vervet monkeys: "The behavior of all parties has been genetically selected by its contribution to the survival of vervet monkeys." He says this in the face of the evidence reported by Seyfarth, Cheney, and Marler (1980) of training (or, if you like, operant conditioning) of the young vervets by the adults. He then compares the case of the vervets to human language use:

> Speakers of English have been conditioned by a verbal community in such a way that when two or more of them are crossing a busy street, the one who sees a danger "emits a call" in response to which the others take appropriate action. There is one call for trucks, another for an open manhole.

But whereas a relatively simple, killjoy, behavioristic account like this *might* turn out to be true of the vervets, we already know perfectly well that it is false of English speakers: "The one who sees a danger 'emits a call.'" Always? No. First, seeing a danger is one (nonintentional) thing; seeing a danger *as a danger* is another thing, and unless one recognizes (actual) dangers as dangers, one is surely not going to emit any danger calls. It would take an intentional characterization to add this obviously necessary restriction. Moreover, we know that our fellow humans are quite up to the nastiness of leading enemies (whom they *want* to hurt) into the paths of trucks, for instance. (As **Terrace** notes, a good question to raise about vervets is whether they are capable of something analogous.) Or, more benignly, humans are up to forgoing the "call" and trying some other act when they *believe* their audience is deaf, to take just one possible case. So the first of Skinner's sentences quoted above must be incorrect. It could be brought a little closer to the truth, no doubt, by inserting a "usually," but the intentionalist or mentalist can do much better: The intentionalist can say when and why the "usual" calls will *not* be "emitted" – *and* when a false danger call might be emitted (otherwise most improbably) in the absence of any danger or even anything seen *as* a danger. In short, cases that at best disappear into the statistical noise of **Skinner**'s and **Rachlin**'s probability claims can be singled out for special predictive and explanatory treatment by the intentionalist, whose attributions of belief and desire and other intentional states provide "hidden variables" to rely on in giving higher probability predictions. (**Millikan** finds a different way to describe the contrast between a behavioristic way of organizing one's data and an intentionalistic way: She points out how the concept of information serves to enlarge the horizons of the scientist describing the relevant conditions and variables.)

The second of **Skinner**'s sentences quoted above reveals another well-known problem: "one call" for trucks? Which is it? "Truck!"? "Look out!"? "Look out for that truck!"? "Back!"? Exercise for undergraduates: Make a list of fifty different English "calls" that might be emitted in this circumstance, and say what they have in common (so we can call them "one call" after all). Exercise for postdocs: Now try to say what they have in common without relying on intentional or semantic terms.

Have others noticed how curiously bland **Skinner**'s assertion style is? Aren't behaviorists, like other scientists, supposed to try for "every" and "always"? By avoiding these quantifiers Skinner forestalls the barrage of counterexamples that would otherwise be hurled at him, while still giving the impression that he is informally advancing general claims. But a more powerful source of superficial plausibility is the subliminal encouragement to the reader to do what comes naturally: Supply the *intentional* interpretation that brings those assertions close to the truth. Perhaps some people find Skinner convincing because they don't realize that they are interpreting what they read with the aid of officially illicit (mentalistic) constructions.

Rachlin's commentary exhibits the same phenomenon. For some – not all – purposes, he says, statements of belief can be effectively replaced by statements about probability. But by the same token, on some occasions statements about belief, such as "I believe that your belief is mistaken," can be effectively replaced by a simple expletive. Neither replacement is a translation or reduction or even an element in such a translation or reduction. So, contrary to the impression given, nothing follows from the claim. In particular, the claim does nothing to erode the generalization that no behaviorist has *ever* succeeded in giving "behavioral criteria" for a mentalistic idiom – or succeeded in living without such idioms. The failure of the behavioristic "reductions," and the reasons for it, have long been familiar to philosophers (Dennett 1969; 1978a; Fodor 1968; 1975; Quine 1960; Taylor 1964). Why, I ask myself, does Rachlin believe otherwise? And why does he believe that his sentence "The parrot says, 'Polly wants a cracker'" exhibits opacity, when it doesn't? (For the standard discussion, see Quine 1960; for my discussion see Dennett 1969, chap. 2.) These and similar "why?" questions baffled me until it dawned on me that these failed attempts of mine to adopt the intentional stance toward Rachlin's commentary were clues leading to the point I was supposed to see: I was asking the wrong sort of questions! Instructed, then, I have changed my tack in a direction Rachlin should (um) be reinforced by. I am now imbued with scientific curiosity about just what sort of history of reinforcement could explain the reading and writing behavior manifested by Rachlin.

Roitblat and **Terrace** usefully explore the analysis of belief attribution to animals, and both claim that "optionality" or being "voluntary" is an important feature.

While there is surely something right and important about this, I have doubts about their formulations. Perhaps I have misunderstood Roitblat, but from his account of optionality it seems to be altogether too relativized to our ignorance at any moment to be a good descriptive term. Thus if some act or response p "normally follows" q but on some occasion occurs in the absence of q, we have it that p has "positive optionality" – but this must be relative to our discernment of interesting values of, or variations on, q. It may be that all previous q's cooccurred with r's, and r is also present on this occasion; relative to r, then, p has no positive optionality – may not be optional at all. But if this is what Roitblat has in mind when he notes the consistency of the intentional stance and scientific determinism, I don't know why he thinks optionality defined thus will be a well-behaved concept. (One further quibble about optionality: Only *novel* deceptive moves would meet Roitblat's test for "positive optionality" – witness the open questions that surround the interpretation of distraction-displaying birds.)

Terrace suggests that one determines the voluntary nature of an act by finding circumstances in which the animal "elects" not to do X. Fine. That's just what the intentional stance is for: describing circumstances in which, given the beliefs and desires inculcated thereby, the organism would find some alternative course of behavior appropriate and "elect" it. Terrace is in fact using the intentional stance without fully acknowledging it. For instance, he offers a "nonintentional rule": "When danger$_i$ is seen and, when within shouting distance of other vervet monkeys, produce alarm call$_i$." If we interpret this *from the vervet's point of view* in effect, as a rule to be followed, it seems nonintentional, though our reliance on point of view commits us to "mentalism." If we don't view it as a rule to be followed, but rather as a regularity observed in monkey behavior, we must reinsert the intentional idioms left out (as in **Skinner**'s example, discussed above): " When something – dangerous or not – is *seen as* a danger$_i$, and when the monkey *believes* it is within shouting distance of other vervet monkeys, it produces alarm call$_i$." Also, I must correct a misapprehension expressed by Terrace; I don't "conclude" that vervet alarm calls are any particular order; I entirely agree that I "must await" the outcome of interesting experiments.

My discussion of *referential opacity* as a mark of intentionality must be counted as expository failure, since **Bennett** and **Harman** declare that it is a red herring at best, while **Rachlin** gets it wrong and **Dawkins** and **Jolly** express puzzlement. My point was not to define or give the essence of intentionality; that is a longish and controversial business, which I have attempted elsewhere. (Those interested in an encyclopedia-style account might consult my entry on intentionality, written with John Haugeland, forthcoming.) My point was simply that one can often uncover covert "mentalism," or reliance on the intentional stance, by spotting referentially opaque contexts, and that the power of such mentalistic locutions depends on their capacity to distinguish between different ways of referring to (thinking of, being about) things. Harman and Bennett are right that there are nonmentalistic opaque contexts.

I drew attention to opacity in order to be able to make the sort of claim exemplified in my replies to **Skinner** and **Terrace**, where the sensitivity to description plays an important theoretical role. **Jolly** introduces a similar case: What is the role of opacity in characterizing the fear of, say, an elephant shrew? Answer: perhaps none. There is, apparently, a phenomenon of pure panic that has no "intentional object," and the same is true of a number of other mental or emotional states. But *if* the fear or emotion of the creature is cognitively complex, we can keep tract of the aspects under which environmental objects can provoke it by relying on opaque construals. Thus Seyfarth qua *rustling-thing-in-the-bush* and not qua *member-of-the-UCLA-faculty* is the object of the creature's terror. This is, in effect, what Jolly notes, in different words.

IV. Stalking the elusive adaptationist. What was the real reason, **Humphrey** wonders, why I tied intentional system theory to adaptationist thinking in biology? Not just because the intentional stance is adaptive, but because there are a wealth of parallels between the intentional stance and adaptationist thinking. The commentaries help bring this out, and I am sure I am not alone in finding the perspectives provided by **Eldredge, Dawkins, Ghiselin,** and **Maynard Smith** very useful in orienting the debate about adaptationism for outsiders. For instance, Maynard Smith gives a good example of such cross-illumination in his remarks on the controversy among psychologists concerning the "matching law" versus "optimal behaviour." And as Dawkins (among others) notes, the virtual unfalsifiability of the two *stances* is unproblematically consistent with the falsifiability of particular attributions and explanations.

The most important parallel, I think, is this: Psychologists can't do their work without the rationality assumptions of the intentional stance, and biologists can't do their work without the optimality assumptions of adaptationist thinking – though some in each field are tempted to deny and denounce the use of these assumptions. Just as confusion and controversy continue to surround the imputation of rationality – as in the use of a "principle of charity" – by philosophers and psychologists attributing intentional states to organisms and people, so there is plenty of talking at cross-purposes among biologists about the role of optimality assumptions. The confusions seem to me to have very similar causes, and very similar effects. Thus **Ghiselin** says "it is an egregious blunder to claim that the study of evolutionary adaptation posits optimality in any interesting or significant way." Indeed it is, but whose blunder is it? As **Maynard Smith** says, 'in using optimisation, we are not trying to confirm (or refute) the hypothesis that animals always optimise." What counts as *positing?* Would Ghiselin agree that Maynard Smith is innocent of this blunder? Then who is left to endure the ridicule that goes with being a Panglossian? Refutation by caricature is a pointless game, however amusing, since any theoretical position, however sound, admits of easy caricature, which in turn is easily "refuted." Thus Ghiselin says the typical Panglossian question is, What is good? But what adaptationist research program is fairly described as asking *that* question? (Cf. **Eldredge.**) Ghiselin proposes to replace the silly question with, "What has happened? The new question does everything we could expect the old one to do, and a lot more besides." This does sound to me like **Skinner**'s familiar claim that the question "What is the history of reinforcement?" is a vast improvement over "What does this person believe, want, intend?" But how much can we actually say in response to this "better question" without a healthy helping of (safe) adaptationist assumptions? The fossil record can certainly be used to answer questions about what, where, and when, but as soon as we turn to *how* (let alone *why*), it seems to me we have to rely on adaptationist assumptions. It may seem as if a scrupulously nonadaptationist science can tell us everything we want to know about "what has happened," but that, I think, is an illusion – like the illusion of plausibility in Skinner's commentary I remarked on above.

Consider **Eldredge**'s example of Fisher's (1975) research on horseshoe crab swimming speed. I know this research only through Eldredge's account, and it does seem that it answers a "what has happened?" question quite persuasively, but the answer *depends on* a very safe adaptationist assumption about what is good: *Faster is better – within limits.* The conclusion that Jurassic horseshoe crabs swam faster depends on the premise that they would achieve maximal speed, given their shape, by swimming at a certain angle, *and* that they would swim so as to achieve maximal speed. So in addition to Fisher's more daring use of optimality considerations conceded by Eldredge, there is his presumably entirely uncontroversial, indeed tacit, use of optimality considerations in order to get *any purchase at all* on "what happened" 150 million years ago. So I can't see how Eldredge can claim that the notion of adaptation is "naught but conceptual filigree" in Fisher's research.

This can be seen from another angle if we revert to the other side of my coin for a moment and examine **Graham**'s and **Harman**'s suggestion that the Skinnerians are the real tellers of "just so" stories. Graham attributes this claim to me (in Dennett 1978a, chaps. 4, 5), but there my argument was slightly different (see pp. 69–70). My point was that when **Skinner** claimed that the true explanation for some complex and novel item of human behavior (observed "in the wild," not in the laboratory) lies (somewhere) in the history of reinforcement of the subject, he is invoking a worse *virtus dormitiva* than those he is criticizing. Of course something or other about the history accounts for the current behavior. Similarly something or other about the almost totally inaccessible history of mutation and reproduction of a species accounts for its various genetically controlled features. But if one wants to give a better answer to the question "What has happened?" than just "something or other," one is going to have to rely *somewhat* on adaptationist thinking. If **Eldredge** agrees with **Ghiselin** that the new biology can "reject . . . teleology altogether" while asking its historical questions, his own example fails to show it, just as Skinner's appeals to the history of reinforcement fail to show how in fact one can get along without mentalism.

I turn now to **Lewontin,** whose main contention is that I am succumbing to a confusion between adaptation and adaptationism. Another egregious blunder, or perhaps the same one in a slightly different guise. Again, whose blunder is it? Are there any adaptationists? What is adaptationism, according to Lewontin?

Lewontin reminds us of genetic hitchhiking and random genetic drift. Given those two phenomena, he says, it is simply factually incorrect to describe evolution as always being an adaptive or optimizing process. Is *this* the defining error of adaptationism? How could it be? To whom is Lewontin addressing these remarks? He may suppose if he wishes that a philosopher has never heard of genetic hitchhiking or random genetic drift, but surely the biologists he is supposedly criticizing are not in need of this textbook review. He says as much. So they must disagree about the implications of these recognized facts. I think I can see why. The claim Lewontin calls factually incorrect is actually subtly equivocal. There is a "grain problem" here. If one looks closely enough at evolution, one sees plenty of important perturbations and exceptions to adaptation; lots of noise, some of which gets amplified by procreation. So evolution is not always adaptive. Q.E.D. But if we step back a bit, we can say, without denying what has just been granted, that evolution is always a *noisy* adaptive process, always adaptive in the long run. Is one an adaptationist if one chooses to look at the whole process that way? If so, is it a mistake? Perhaps Lewontin would say that it is always a mistake to look less closely at evolutionary processes than one can, but that is an implausible methodological canon. It is often useful to abstract – to say (rigorous, falsifiable) things about the forest and let someone else worry about the trees.

It would surely be a mistake to assume that evolution was adaptive *all the way down* without any exception, but if that is adaptationism ("a world view that raises a phenomenon to untested universality"), I wonder if there are any adaptationists today. But **Lewontin** gives another, different characterization of adaptationism that looks more realistic: There is a body of evidence that is "simply sidestepped by Panglossian adaptationists who find it inconvenient." Is there anything wrong with that? Once again, let's see how this issue looks on the other side of my coin. One often hears it said by neuroscientists that there is a mass of data and theory about the fine structure and operation of the brain that no one denies, but which is simply sidestepped by the cognitivists – the "top-down" mentalists – who find it inconvenient. True. So what? It's not a bad idea at all, although of course it can be carried to excess, like anything else. As **Beatty** says, in his useful and irenic reinterpretation of the controversy, Gould and Lewontin's urgings make much more sense as a call for a multiplicity of research programs each concentrating on a way of making progress, than as a condemnation of any of the special interest groups of such a pluralistic society.

Who is confusing adaptation with adaptationism? At one point **Lewontin** says my "rhetorical flummery" is to suggest he and Gould have as a hidden agenda the "extirpation root and branch of adaptation"; *what I said* was "extirpation root and branch of adaptationism" – and I was explicitly denying that this was their intent. Here we get to the heart of the matter: a persistent failure of communication that our prior correspondence (to which Lewontin alludes) failed to correct, and which I abandoned once it began to take on the tone of Lewontin's commentary. If I was confusing adaptation with adaptationism, it was because I was also mistakenly supposing that adaptationism was a position actually held by someone. I thought an adaptationist was one who favored and even concentrated on adaptationist reasoning, not a person who was silly enough to raise a phenomenon to untested universality. I defended the former; if Gould and Lewontin are in fact out to extirpate the latter variety root and branch, they needn't try so hard; so far as I know, there aren't any anymore. So I'll admit my blunder if Lewontin will admit to shooting at his own shadow.

What about Harvard conservatism? In Dennett (1971) I proposed an economic metaphor: Indulging in intentional discourse was taking out an intelligence loan, which ultimately had to be repaid. Quine and Skinner, I pointed out, were, in terms of this metaphor, rock-ribbed New England fiscal conservatives who disapprove of deficit spending, who caution everyone against taking out any loans of this sort. In Gould and Lewontin's attack I see the same puritanical disapproval of this practice of helping oneself to adaptationist assumptions. The adaptationist agrees that the loans must all be paid back. Consider, for instance, how Dawkins scrupulously pauses, again and again, in *The Selfish Gene* (1976) to show precisely what the cash value of his selfishness talk is. Nevertheless, some people – a certain *sort* of conservative – deeply disapprove of this way of doing business, whether in philosophy, psychology, or biology. It is probably just an amusing coincidence, however, that Quine, Skinner, Lewontin, and Gould are all at Harvard.

Finally, what are we to make of the uncharacteristic, apparently unaccountable lapses in **Lewontin**'s commentary? For instance:

1. the unsupported charges – not a single citation – of "elementary errors" on my part.
2. the complete misreading of my friendly italics: "Gould and Lewontin *seem* to be championing . . . a scrupulously nonadaptationist, historic-architectural answer." What I meant was that although this is what they seem to many of their supporters to be doing, in fact they are espousing pluralism, as they insist and I acknowledge. Has Lewontin forgotten that he is *not* a scrupulous nonadaptationist, but rather an open-minded pluralist?

I think the explanation of this disappointing phenomenon is straightforward. I try to practice what I preach, and the target article was itself a Sherlock Holmes experiment of sorts. Noting that **Lewontin** is apparently a proficient utterer of a certain sort of speech act – "abusing" adaptationists, as he puts it – I asked myself whether he was also proficient in the audience role for such acts. More poetically, could he take a joke?

Apparently not. One whiff of Skinneric acid is enough to overpower his good sense and trigger a distraction display, complete with a quite affecting broken-left-wing dance (brandishing the infamous list of never-to-be-revealed names). But a single demoting experiment is not conclusive, as Lloyd Morgan realized. It just goes to show: Nobody's perfect.

References

Barash, D. P. (1976) Male response to apparent female adultery in the mountain bluebird: An evolutionary interpretation. *American Naturalist* 110:1097–1101. [taDCD]

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

Bennett, J. (1964) *Rationality*. Routledge and Kegan Paul. [JBen]

(1976) *Linguistic behaviour*. Cambridge University Press. [JBen, tarDCD, AR]

Boden, M. (1981) The case for a cognitive biology. In: *Minds and mechanisms: Philosophical psychology and computational models*. Cornell University Press. [taDCD]

Boring, E. G. (1950) *A history of experimental psychology*. 2d ed. Appleton-Century-Crofts. [MSS]

Cain, A. J. (1964) The perfection of animals. *Viewpoints in Biology* 3:37–63. [taDCD]

Cargile, J. (1970) A note on "iterated knowings." *Analysis* 30:151–55. [taDCD]

Charnov, E. L. (1982) *The theory of sex allocation*. Princeton University Press. [MG]

Cherniak, C. (1981) Minimal rationality. *Mind* 99:161–83. [taDCD]

Churchland, P. M. (1981) Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78:67–90. [PSC, ACD, rDCD, AR]

Churchland, P. S. (1980a) Language, thought, and information processing. *Nous* 14:147–69. [PSC]

(1980b) A perspective on mind-brain research. *Journal of Philosophy* 78:185–207. [PSC]

Churchland, P. S. & Churchland P. M. (1983) Stalking the wild epistemic engine. *Nous*, in press. [PSC]

Clark, S. R. L. (1982) *The nature of the beast*. Oxford University Press. [RD]

Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. (1982) *Red deer: Behavior and ecology of two sexes*. University of Chicago Press. [RD]

Clutton-Brock, T. H., & Harvey, P. H. (1979) Comparison and adaptation. *Proceedings of the Royal Society of London, B*, 205:547–65. [RD]

Cohen, L. J. (1981) Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences* 4:317–70. [rDCD]

Danto, A. C. (1983) Towards a theory of retentive materialism. In: *How many questions: Essays in honor of Sidney Morgenbesser*, ed. L. Cauman, I. Levi, C. Parsons & R. Schwartz. Hackett. [ACD]

Darwin, C. R. (1872) *The expression of the emotions in man and animals*. John Murray. [MG]

Davis, J. D. & Wirtshafter, D. (1978) Set-points or settling points for body weight? A reply to Mrosovsky and Powley. *Behavioural Biology* 24:405–11. [DMcF]

Dawkins, M. (1980) *Animal suffering*. Chapman & Hall. [RD]

Dawkins, R. (1976) *The selfish gene*. Oxford University Press. [tarDCD, DL]

(1980) Good strategy or evolutionarily stable strategy? In: *Sociobiology: Beyond nature nurture?*, ed. G. W. Barlow & J. Silverberg. A.A.A.S. Selected Symposium. Westview Press. [taDCD]

(1982) *The extended phenotype*. W. H. Freeman [RD]

Dawkins, R. & Krebs, J. R. (1978) Animal signals: Information or manipulation? In *Behavioural ecology*, ed. J. R. Krebs & N. B. Davies, pp. 289–309. Blackwell Scientific Publications. [RD]

Dennett, D. (1969). *Content and consciousness*. Humanities Press. [rDCD, DL, AR]

(1971) Intentional systems. *Journal of Philosophy* 68:87–106. Repr. in *Dennett* 1978a. [tarDCD, DL, RGM]

(1975) Brain writing and mind reading. In: *Language, mind, and knowledge*, Minnesota Studies in the Philosophy of Science vol. 7, ed. K. Gunderson. Repr. in *Dennett* 1978a. [DL]

(1976) "Conditions of personhood." In: *The identities of persons*, ed. A. O. Rorty. University of California Press. Repr. in *Dennett* 1978a. [taDCD]

(1978a) *Brainstorms*. Bradford/MIT Press. [PSC, tarDCD, GC, DL]

(1978b) Reply to Arbib and Gunderson. In: *Brainstorms*. Bradford/MIT Press. [DL]

(1978c) Why not the whole iguana? *Behavioral and Brain Sciences* 1:103–4. [taDCD, RGM]

(1979) On the absence of phenomenology. In: *Body, mind, and method*, ed. D. F. Gustafson & B. L. Tapscott, pp. 93–113. D. Reidel Publ. [DL]

(1980) Passing the buck to biology. *Behavioral and Brain Sciences* 3:19. [rDCD]

(1981a) Making sense of ourselves. *Philosophical Topics* 12: 63–81. [taDCD]

(1981b) Three kinds of intentional psychology. In: *Reductionism, time and reality*, ed. R. Healey. Cambridge University Press. [PSC, tarDCD, RGM]

(1981c) True believers: The intentional strategy and why it works. In: *Scientific explanation*, ed. A. Heath. Oxford University Press. [tarDCD, DL, RGM]

(1982) How to study consciousness empirically: Or, nothing comes to mind. *Synthese* 53:159–80. [tarDCD]

(1983) Styles of mental representation. *Proceedings of the Aristotelian Society*, in press. [rDCD]

Dennett, D. C. & Haugeland, J. (forthcoming) Intentionality. In: *The Oxford companion to the mind*, ed. R. Gregory. Oxford University Press. [rDCD]

Dobzhansky, T. (1956) What is an adaptive trait? *American Naturalist* 90:337–47. [taDCD]

Dover Wilson, J. (1951) *What happens in Hamlet*. 3rd ed. Cambridge University Press. [taDCD]

Dretske, F. (1981) *Knowledge and the flow of information*. Bradford/MIT Press. [taDCD]

Duhem, P. (1906) *La théorie physique: Son object et sa structure*. Chevalier et Riviere. [GH]

Eisenberg, J. F. (1981) *The mammalian radiations*. Athlone Press. [RD]

Fisher, D. (1975) Swimming and burrowing in *Limulus* and *Mesolimulus*. *Fossils and Strata* 4:281–90. [rDCD, NE]

Fodor, J. A. (1968) *Psychological explanation*. Random House. [rDCD]

(1975) *The language of thought*. Crowell. [PSC, rDCD]

(1980) Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences* 3:63–109. [PSC]

von Frisch, K. (1967) *The dance language and orientations of bees*. Translated by L. E. Chadwick. Belknap Press of Harvard University Press. [HST]

Gardner, R. A. & Gardner, B. T. (1978) Comparative psychology and language acquisition. *Annals of the New York Academy of Sciences* 309:37–76. [MSS]

Ghiselin, M. T. (1969) *The triumph of the Darwinian method*. University of California Press. [MG]

(1974) *The economy of nature and the evolution of sex*. University of California Press. [MG]

(1981) Categories, life, and thinking. *Behavioral and Brain Sciences* 4:269–83. [MG]

(1982) On the mechanisms of cultural evolution, and the evolution of language and the common law. *Behavioral and Brain Sciences* 5:11. [MG]

Gould, J. L. & Gould, C. G. (1982) The insect mind: Physics or metaphysics? In: *Animal mind–human mind*, ed. D. R. Griffin. Dahlem Workshop. Springer-Verlag. [tarDCD]

Gould, S. J. (1977) *Ever since Darwin*. W. W. Norton & Co. [taDCD]

(1980) *The panda's thumb: More reflections in natural history*. W. W. Norton & Co. [MG]

(1981) *The mismeasure of man*. Norton. [taDCD]

Gould, S. J. & Lewontin, R. (1979) The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society* (London) B205:581–98. [RD, taDCD, NE, RCL, AR, MSS]

Grice, H. P. (1957) Meaning. *Philosophical Review* 66:377–88. [taDCD, GH, JH, HST]

(1969) Utterer's meaning and intentions. *Philosophical Review* 78:147–77. [taDCD, JH, HST]

Griffin, D. R. (1978) Prospects for a cognitive ethology. *Behavioral and Brain Sciences* 1:527–38. [DRG, HST]

(1981) *The question of animal awareness*. 2d ed. Rockefeller University Press. [RD, DRG, CAR]

Harcourt, A. H., Harvey, P. H., Larson, S. G. & Short, R. V. (1981) Testis weight, body weight and breeding system in primates. *Nature* 293:55. [RD]

Harman, G. (1973) *Thought*. Princeton University Press. [rDCD]

(1974) Review of *Meaning*, by Stephen Schiffer. *Journal of Philosophy* 71:224–29. [GH]

Harvey, P. H., Clutton-Brock, T. H. & Mace, G. (1980) Brain size and ecology in small mammals and primates. *Proceedings of the National Academy of Sciences* 77:4387–89. [RD]

Heil, J. (1982) Speechless brutes. *Philosophy and Phenomenological Research* 42:400–406. [JH]

(1983) *Perception and cognition*. University of California Press. [JH]

Hempel, C. G. (1950) Problems and changes in the empiricist criterion of meaning. *Revue Internationale de Philosophie* 11:41–63. [GH]

Hinde, R. A. (1970) *Animal behaviour*. McGraw-Hill. [DMcF]

von Holst, E. & Mittelstaedt, H. (1950) Das Reafferenzprinzip. *Naturwissenschaften* 37:464–76. [DMcF]

Hulse, S. H., Fowler, H. & Honig, W. K. (1978) *Cognitive processes in animal behavior*. Lawrence Erlbaum Associates. [HST]

Humphrey, N. K. (1976) The social function of intellect. In: *Growing points in ethology*, ed. P. P. G. Bateson & R. A. Hinde. Cambridge University Press. [taDCD]

(1979) Nature's psychologists. In: *Consciousness and the physical world*, ed. B. D. Josephson & V. S. Ramachandran, pp. 57–75. Pergamon Press. [NH]

(1980) Nature's psychologists. In: *Consciousness and the physical world*, ed. B. D. Josephson & V. S. Ramachandran, pp. 57–75. Pergamon Press. [GG]

(1982) Consciousness: a Just-So story. *New Scientist* 95:474–77. [NH]

Jerison, H. J. (1973) *Evolution of the brain and intelligence*. Academic Press. [RD]

Johnston, T. D. (1981) Contrasting approaches to a theory of learning. *Behavioral and Brain Sciences* 4:125–73. [taDCD]

Kahneman, D. (unpublished) Some remarks on the computer metaphor. [taDCD]

Kalman, R. E. (1963) Mathematical description of linear dynamical systems. *Journal of the Society for Industrial and Applied Mathematics Control* Series A.1. 152–92. [DMcF]

Köhler, W. (1925) *The mentality of apes*. Liveright. [EWM]

(1947) *Gestalt psychology*. Liveright. [EWM]

Kummer, H. (1982) Social knowledge in free-ranging primates. In: *Animal mind–human mind*, ed. D. R. Griffin, pp. 113–30. Dahlem Workshop. Springer-Verlag. [AJ]

Kyburg, H. E., Jr. (1983) Rational belief. *Behavioral and Brain Sciences* 6:231–73. [rDCD]

Lewin, R. (1980) Evolutionary theory under fire. *Science* 210:881–87. [taDCD]

Lewis, D. (1974) Radical interpretation. *Synthese* 23:331–44. [taDCD]

Lewontin, R. (1961) Evolution and the theory of games. *Journal of Theoretical Biology* 1:328–403. [taDCD]

(1972) Testing the theory of natural selection. *Nature* 236:181–82. [JBea]

(1978a) Adaptation. *Scientific American* 213–30. [taDCD]

(1978b) Fitness, survival and optimality. In: *Analysis of ecological systems*, ed. D. H. Horn, R. Mitchell & G. R. Stairs. Ohio State University Press. [taDCD]

(1979) Sociobiology as an adaptationist paradigm. *Behavioral Science* 24:5–14. [RD, taDCD]

(1981) The inferiority complex. *New York Review of Books*, October 22, pp. 12–16. [taDCD]

Lloyd, M. & Dybas, H. S. (1966) The periodical cicada problem. *Evolution* 20: 132–49; 466–505. [taDCD]

Lorenz, K. Z. (1971) *Studies in animal and human behavior*. Harvard University Press. [EWM]

McFarland, D. J. (1971) *Feedback mechanisms in animal behaviour*. Academic Press. [DMcF]

McFarland, D. & Houston, A. (1981) *Quantitative ethology*. Pitman Books. [rDCD, DMcF]

Martin, R. D. (1981) Relative brain size and basal metabolic rate in terrestrial vertebrates. *Nature* 293:57–60. [RD]

Maynard Smith, J. (1972) *On evolution*. Edinburgh University Press. [taDCD]

(1974) The theory of games and the evolution of animal conflict. *Journal of Theoretical Biology* 49:209–21. [taDCD]

(1978) Optimization theory in evolution. *Annual Review of Ecology and Systematics* 9:31–56. [taDCD]

Mayr, E. (1982) *The growth of biological thought*. Harvard University Press. [EWM]

Menzel, E. W. (1969) Naturalistic and experimental approaches to primate behavior. In: *Naturalistic viewpoints in psychological research*, ed. E. Willems & H. Raush. Holt, Rinehart and Winston. [EWM]

(1971) Communication about the environment in a group of young chimpanzees. *Folia Primatologica* 15:220–32. [EWM]

(1974) A group of young chimpanzees in a one-acre field. In: *Behavior of nonhuman primates*, vol. 5, ed. A. M. Schrier & F. Stollnitz. Academic Press. [EWM]

(1979) General discussion of the methodological problems involved in the study of social interaction. In: *Social interaction analysis: Methodological issues*, ed. M. Lamb & G. Stephenson. University of Wisconsin Press. [EWM]

Menzel, E. W. & Johnson, M. K. (1976) Communication and cognitive organization in humans and other animals. *Annals of the New York Academy of Sciences* 280:131–42. [EWM]

(1978) Should cognitive concepts be defended or assumed? *Behavioral and Brain Sciences* 4:586–87. [EWM]

Midgley, M. (1979) Gene juggling. *Philosophy* 54:439– 58. [RD]

Miller, G., Galanter, E. & Pribram, K. (1960) *Plans and the structure of behavior*. Holt, Rinehart and Winston. [EWM]

Millikan, R. G. (forthcoming) *Language, thought, and other biological categories*. Bradford/MIT Press. [rDCD]

Morgan, C. L. (1894) *An introduction to comparative psychology*. Walter Scott. [JH, DMcF]

(1900) *Animal behaviour*. Walter Scott. [DMcF]

(1909) *An introduction to comparative psychology*. 2d. ed. Walter Scott. [MG]

Mrosovsky, N. & Powley, T. L. (1977) Set points for body weight and fat. *Behavioural Biology* 20: 205–25. [DMcF]

Newell, A. (1982) The knowledge level. 1980 presidential address, American Association for Artificial Intelligence. *Artificial Intelligence* 18:87–127. [tarDCD]

Nozick, R. (1974) *Anarchy, state, and utopia*. Basic Books. [GH]

(1981) *Philosophical explanations*. Harvard University Press. [taDCD]

Oster, G. F. & Wilson, E. O. (1978) *Caste and ecology in the social insects.* Princeton University Press. [RD, taDCD]
Patterson, F. (1978) The gestures of a gorilla: Language acquisition in another pongid. *Brain and Language* 5:72–97. [MSS]
Patterson, F. & Linden, E. (1981) *The education of Koko.* Holt, Rinehart and Winston. [HST]
Premack, D. & Woodruff, G. (1978) Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1:515–26. [taDCD, AJ]
Quine, W. V. O. (1951) Two dogmas of empiricism. *Philosophical Review* 60:20–43. [GH]
(1960) *Word and object.* MIT Press. [tarDCD]
Rachlin, H., Battalio, R., Kagel, J. & Green, L. (1981) Maximization theory in behavioral psychology. *Behavioral and Brain Sciences* 4:371–418. [HR]
Reddinguis, J. (1980) Control theory and the dynamics of body weight. *Physiology and Behaviour* 24:27–32. [DMcF]
Ridley, M. (1983) *The comparative method and adaptations for mating.* Oxford University Press. [RD]
Ristau, C. A. (1983) Language, cognition and awareness in animals? In: *The use of animals in biomedical research,* ed. J. Sechzer. New York Academy of Sciences. [CAR]
Roitblat, H. L. (1982) The meaning of representation in animal memory. *Behavioral and Brain Sciences* 5:352–406. [taDCD, HLR, HST]
Romanes, G. J. (1882) *Animal intelligence.* Kegan Paul, Trench. [DMcF, MSS]
Rosenberg, A. (1980) *Sociobiology and the preemption of social science.* Johns Hopkins University Press. [taDCD]
Russell, B. (1905) On denoting. *Mind,* pp. 479–93. Repr. in Russell, B. (1958) *Logic and knowledge,* Allen & Unwin. [taDCD]
(1919) *Introduction to mathematical philosophy.* (1971 Reprint). Touchstone Books. [rDCD]
Sarkar, H. (1982) A theory of group rationality. *Studies in History and Philosophy of Science* 13:55–72. [JBea]
Savage-Rumbaugh, S., Rumbaugh, D. M. & Boysen, S. (1978) Linguistically mediated tool use and exchange by chimpanzees (*Pan troglodytes*). *Behavioral and Brain Sciences* 1:539–54. [JBen, taDCD]
Schank, R. C. (1976) Research at Yale in natural language processing. Research report 84, Yale University Department of Computer Science. [taDCD]
Schwartz, B. & Lacey, H. (1982) *Behaviorism, science, and human nature.* W. W. Norton & Co. [GC]
Seidenberg, M. S. & Petitto, L. A. (1979) Signing behavior in apes: A critical review. *Cognition* 7:177–215. [MSS]
(1981) Ape signing: Problems of method and interpretations. *Annals of the New York Academy of Sciences* 364:115–29. [MSS]
Seyfarth, R., Cheney, D. L. & Marler, P. (1980) Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science* 210:801–3. [tarDCD, HST]
Shannon, C. (1949) *The mathematical theory of communication* (with an introductory essay by Warren Weaver). University of Illinois Press. [taDCD]
Sheffield, F. D. (1965) Relation between classical conditioning and instrumental learning. In: *Classical conditioning,* ed. W. F. Prokasy. Appleton-Century-Crofts. [HST]
Shwayder, D. (1965) *The stratification of behaviour.* Roultedge and Kegan Paul. [rDCD]
Simmons, K. E. L. (1952) The nature of the predator reactions of breeding birds. *Behaviour* 4:101–76. [taDCD]
Simon, H. (1957) *Models of man.* Wiley. [taDCD]
Skinner, B. F. (1931) The concept of the reflex in the description of behavior. *Journal of General Psychology* 5:427–58. [HST]
(1957) *Verbal behavior.* Appleton-Century-Crofts. [BFS]
(1964) Behaviorism at fifty. In: *Behaviorism and phenomenology: Contrasting bases for modern psychology,* ed. T. W. Wann. University of Chicago Press. [taDCD]
(1971) *Beyond freedom and dignity.* Knopf. [taDCD]
(1981) Selection by consequences. *Science* 213:501–4. [BFS]
Skutch, A. F. (1976) *Parent birds and their young.* University of Texas Press. [taDCD]
Small, W. S. (1901) Experimental study of the mental processes of the rat. *American Journal of Psychology* 12:218–20. [HR]
Sommerhoff, G. (1950) *Analytical biology.* Oxford University Press. [EWM]
Sordahl, T. A. (1981) Sleight of wing. *Natural History* 90:43–49. [taDCD]
Stabler, E. P., Jr. (1983) How are grammars represented? *Behavioral and Brain Sciences* 6: 391–421. [rDCD]
Stafford, S. (unpublished) The origins of the intentional stance. Tufts University Working Paper in Cognitive Science. [rDCD]
Stich, S. P. (1982) On the ascription of content. In: *Thought and object,* ed. A. Woodfield, pp. 153–206. Clarendon Press. [PSC]
Süffert, F. (1932) Phänomene visueller Anpassung. *Zeitschrift für Morphologie und Ökologie der Tiere* 26:147–316. [MG]
Sulloway, F. (1979) *Freud, biologist of the mind.* Basic Books. [EWM]
Taylor, C. (1964) *The explanation of behaviour.* Routledge and Kegan Paul. [JBen, rDCD]
Terrace, H. S. (1982a) Can animals think? *New Society* 4:339–42. [HST]
(1982b) Why Koko can't talk. *Sciences* 22:8–10. [HST]
(1983) Animal cognition. In: *Animal cognition,* ed. H. L. Roitblat, T. G. Bever & H. S. Terrace. Lawrence Erlbaum Associates. [HST]
Tinbergen, N. (1965) Behavior and natural selection. In: *Ideas in modern biology,* ed. J. A. Moore, pp. 519–42. Natural History Press. [RD]
Toates, F. M. (1980) *Animal behaviour: A systems approach.* J. Wiley & Sons. [DMcF]
Tolman, E. C. (1932) *Purposive behavior in animals and men.* New York: Appleton-Century-Crofts. Repr. University of California Press, 1949. [HLR]
(1951) *Behavior and psychological man.* University of California Press. [EWM]
(1959) Principles of purposive behavior. In: *Psychology: A study of a science,* vol. 2, ed. S. Koch, pp. 92–157. McGraw-Hill. [HLR]
Trivers, R. L. (1971) The evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35–57. [taDCD]
Wiener, N. (1946) *Cybernetics.* MIT Press. [EWM]
Williams, D. R. & Williams, H. (1969) Auto-maintenance in the pigeon: Sustained pecking despite contingent nonreinforcement. *Journal of the Experimental Analysis of Behavior* 12:511–20. [HST]
Williams, G. C. (1966) *Adaptation and natural selection: A critique of some current evolutionary thought.* Princeton University Press. [MG]
Wilson, E. O. (1975) *Sociobiology: The new synthesis.* Harvard University Press. [taDCD]
Wilson, E. O., Durlach, N. I. & Roth, L. M. (1958) Chemical releasers of necrophoric behavior in ants. *Psyche* 65:108–14. [taDCD]
Wirtshafter, D. & Davis, J. D. (1977) Set points, settling points, and the control of body weight. *Physiology and Behaviour* 19:75–78. [DMcF]
Wittgenstein, L. (1958) *Philosophical investigations.* Blackwell. [rDCD]
Woodruff, G. & Premack, D. (1979) Intentional communication in the chimpanzee: The development of deception. *Cognition* 7:333–62. [taDCD]