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

1. 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).

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 instru-

mentalism 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 certianly 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, GG, 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. [GG]

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]