

Chapter 2

**DISAMBIGUATING ANTHROPOMORPHISM: AN
INTERDISCIPLINARY REVIEW**

John Andrew Fisher

*Department of Philosophy
University of Colorado
Boulder, Colorado 80309*

I. ABSTRACT

I claim that anthropomorphism is neither a well-defined nor clearly fallacious fallacy or mistake. I argue that the casual and unexamined use of "anthropomorphism" as a serious term of criticism is untenable. I begin by reviewing the history of the concept and how it has come to have its present negative connotations while at the same time escaping any serious analysis. I ask what is the logical characterization of the fallacy that merits this status. I reject simplistic characterizations that assume that terms that apply to humans cannot apply to animals. I then propose a taxonomy of types of anthropomorphism and argue that only one type, categorical anthropomorphism, justifies the rhetoric associated with anthropomorphism. I explore and reject philosophical arguments for claiming that this sort of anthropomorphism is a general danger.

II. INTRODUCTION

It is widely believed that we should avoid anthropomorphism in our thinking about animals. In spite of this consensus, it is difficult to find any explicit discussion, let alone adequate analysis, of what anthropomorphism is. The use of the term "anthropomorphism" by scientists and philosophers is often so casual as to almost suggest that it is a term of ideological abuse, rather like political or religious terms ("communist" or "counterrevolutionary") that need no explanation or defense when used in criticism.

Anthropomorphism is generally treated by thinkers as an obvious mistake or fallacy of some sort. It is a term of criticism regularly applied to those who support animal thought and animal rights. Its use is so deeply entrenched that even those thinkers who favor animal rights try to avoid being accused of it. Annette Baier (1985), for example, feels obliged to say, "I see nothing at all anthropomorphic or in any other way absurd in saying that one may 'break faith with' an animal, exploit its trust, disappoint expectations one has encouraged it to have" (p. 150, my emphasis). And Mary Midgley (1978), who is also sympathetic to claims of animal mentality, asserts, "There is nothing anthropomorphic in speaking of the motivation of animals. It is anthropomorphic to call the lion the King of Beasts, but not to talk of him as moved, now by fear, now by curiosity, now by territorial anger" (p. 106).

Philosophers and scientists often approach anthropomorphism as an obstacle to be avoided or as a general problem to be overcome by those who wish to attribute cognitive or emotional states to nonhuman animals. Thus Donald Davidson (1975) claims that "Attributions of intentions and beliefs to animals smack of anthropomorphism" (p. 7). Contrary to this vague consensus, I will argue that there is a considerable amount of confusion about anthropomorphism. Is anthropomorphism: (1) a fallacious type of inference? (2) a bad explanation? (3) a false set of beliefs, or something else? If any one, or some combination, of these concepts, how exactly is it to be characterized? And under what conditions is a charge of anthropomorphism valid? What theory, if any, lies behind the charge of anthropomorphism? These are the questions I shall address.

I should make clear at the beginning that my subject is *not* going to be the issue much discussed by philosophers of whether there is a difference between humans and other animals. Of course there is; each species is unique. Moreover, the differences can be quite substantial. I will not question, for example, the widely debated assumption that only humans have significant language use. So lines of demarcation *can* be drawn between humans and other animals. No one should bother debating the question of whether other animals are just like humans.

I am therefore not going to investigate whether humans are different from or superior to other animals—let's suppose for the sake of argument that we are—but rather the question of whether when we apply, as we commonly do in everyday life, terms to nonhuman animals that we apply to humans, *univocally* as it seems, we commit some sort of fallacy, well articulated under the rubric *anthropomorphism*. The commonplace idea that anthropomorphism occurs when we see animals in human terms or apply human terms to animals is not much help here. To say that we are different or unique cannot be to deny that humans and other animals share many characteristics. The question that needs to be answered is: Which are the human terms that do not apply to animals?

Since there are many contexts in which the notion of anthropomorphism

comes up, let me specify the scope and target of our inquiry. I will be primarily concerned with applications of the concept of anthropomorphism to explanations of animal behavior, as opposed to, say, its use in art history and criticism. In the animal context it is a term of negative criticism, and this contrasts with writing in the arts where "anthropomorphic" is a descriptive term, characterizing a figure or form, implying no mistake or negative criticism. The specific target of my inquiry will be appeals to the concept of anthropomorphism made by philosophers and scientists as a criticism of attempts to produce literal understandings or explanations of animal behavior. The most general description of this "mistake" is: relating to animals as if they were very similar to us. (This formulation is meant to include our emotional and sympathetic responses and actions, as well as explanations.)

My claim will be that this mistake or fallacy is neither well-defined nor clearly fallacious. As a term of logical criticism—and I will argue this is its most central function in the relevant literature—it often fails. I will try to show that there is much to sort out before one can pin down occurrences of anthropomorphism. There are many different conceptions of anthropomorphism and the common ones do not support their common rhetorical use. To substantiate these claims I will sketch a taxonomy of different cases and/or types of anthropomorphism.

Anthropomorphism turns out to be a very complex as well as ill-defined topic. If nothing else, it should be obvious from the actual complexity of the concept that its casual and unexamined use as a serious term of criticism is simply untenable. The charge of anthropomorphism oversimplifies a complex issue—animal consciousness—and it tries to inhibit consideration of positions that ought to be evaluated in a more open-minded and empirical manner.

III. A RHETORICAL HISTORY

It would be a mistake to ignore the historical roots of the concept of anthropomorphism. The unique way the charge of anthropomorphism functions as a negative criticism that needs no explanation itself calls for an explanation. We can understand this function better by briefly canvassing the history of the concept.

"Anthropomorphism" was originally a term used in the theological context for views of God that characterize Him in literal human terms. In particular, to view God as a person of sorts, with indefinitely amplified human powers and characteristics (loving, just, knowing) can be labeled anthropomorphic. That "anthropomorphism" in this sense carries a negative connotation can be seen from Paul Edwards (1967) definition: "Let us, *without implying anything derog-*

atory, refer to the belief that predicates can be applied literally to God as the 'anthropomorphic' conception of God and to the belief that predicates can be applied analogically to God as the 'metaphysical' conception of God" (p. 186, my emphasis). As Edwards tells us, "Among professional philosophers, belief in the metaphysical God has been much more common than belief in an anthropomorphic God. This metaphysical position is at least as old as Aquinas. . . ." (p. 186). Given this background, it is not hard to see why in the Western intellectual tradition "anthropomorphism" came to be a derogatory expression. [It is, I suggest, no accident that, as P. J. Asquith tells us, Japanese primatologists are singularly unconcerned about issues of anthropomorphism in their studies of primates. Cultural (and religious) history cannot be ignored in explaining this fact (see Asquith, 1986).] Anyone can understand a God in personal terms, but only the educated can understand the metaphysical God. Moreover, the philosophers had powerful arguments against understanding God anthropomorphically. In addition, pagan religions tended to be anthropomorphic. So Christian doctrine was distanced from primitive forms of understanding, whether exemplified by uneducated belief or by pagan religions. Anthropomorphism in this context is perceived to be a vulgar mistake.

This is one strand. Others wind through the history of science. Here, as in theology, a more primitive and uneducated form of thinking is to be replaced by a more sophisticated and educated style of thought. In discussing the rise of modern science, for example, Ernst Cassirer (1950) associates Spinoza's rejection of Aristotelian causal explanations with the issue of anthropomorphism:

When Aristotle said that God moves all things the same way as does a "beloved object" . . . his very metaphor betrayed the highly dangerous origin of the Aristotelian concepts of form and purpose. The usage of these concepts in metaphysics and science revealed itself to be simply a case of anthropomorphism. Until this is destroyed there can be no really universal interpretation of nature. (p. 123)

In this context "anthropomorphism" refers to views of material nature, in particular causal relations, as at bottom conforming to human experience of causal action on the world. In the more particular domain of biology Cassirer again notes the rejection of anthropomorphism:

It was this elimination and overcoming of all teleological considerations that Haeckel believed to be Darwin's greatest act of emancipation, one comparable with the achievement of Copernicus. Now for the first time there was feasible a treatment of nature from which all the anthropomorphic features had been stripped away. (p. 162)

"Anthropomorphism" is here used for a tendency of thought that must be eliminated from our explanations and theories because it embodies outmoded scientific approaches. We see that Cassirer associates anthropomorphism with the trends of thought that the rise of modern science had to overcome. It is not surprising that traditionally anthropomorphism does not stand for an intellectually respectable position to be taken seriously, but rather a well-known and

unsophisticated fallacy or heresy. To go back to anthropomorphic ways of understanding the world would be like going back on the Copernican revolution, going back to the Dark Ages.

The history rehearsed so far associates the concept of anthropomorphism with outmoded and refuted theories. There is another important strand of the history of science that helped to license the extension of the negative epithet, "anthropomorphic," to the specific context of our understanding of animals. That strand involves the methodological prohibitions sponsored by positivism.

The story of the rise, and now fall, of positivism is both too complex and too familiar to bear telling here. What is relevant for our topic is the antirealist polemic so common to the era of positivism and operationalism, philosophies that dominated concepts of scientific methodology in the first half of the twentieth century. Positivism favored a very austere scientific methodology based on an unwillingness to postulate, in a realist frame of mind, the existence of unobservable entities; that is, genuinely explanatory entities not reducible to observational manifestations. Such strictures derived from a commitment by many positivists to phenomenalism, the view that physical reality can be reduced to the direct objects of perception. Behaviorism, the refusal to countenance appeal to mental states in explanations of human behavior, is the familiar consequence of positivism in psychology. References to mental states are meaningless or, at any rate, not acceptable in rigorous procedures of explanation. Even after such strictures have been abandoned in explanations of human behavior, it is natural to retain them for explanations of other animals. Contrary to behaviorist dogma, it was always widely felt that human mental states *are* observable: we are aware of our own, and we can *ask* others what they think and feel. But nonhuman animals elude this escape from positivistic strictures: we cannot *verify* animal mental states by any direct or semidirect observation; so if we are not to postulate unobservables, we must eschew reference to animal mental states for scientific explanations. [Mary Midgley has extensively explored this epistemological aspect of the rejection of animal mentality in animal studies (see especially Midgley, 1983, pp. 127–131).]

Even though in the last two decades robust realist trends, countenancing postulation of unobservables, have displaced positivism in the philosophy of science, the heritage of positivism's ideas about valid method remains potent [Even thinkers who favor mentalistic explanations of animal behavior sometime try to operate within parsimonious methodologies. For example, Mary Midgley (1978) sounds almost like a behaviorist when she argues, "Members of one species do in fact often succeed in understanding members of another well enough for both prediction and a personal bond. Nothing more is necessary" (p. 348).] Many scientists' ideas about methodology emanate from a time when positivistic scruples dominated. Philosophers too are prone to such scruples. For example, the philosopher Leslie Stevenson (1982) suggests that the "curious

pleasure that most of us seem to derive from such conceptual interpretations of the behaviour of our pets [that they have beliefs, for example] (and of our prelinguistic infants) leads us, at least sometimes, into anthropomorphism which is unjustified by the *strict canons of scientific objectivity*. . ." (p. 17, my emphasis).

Talk of animal mental states is viewed from this perspective as unscientific, a violation of sophisticated methodological scruples about the proper form of explanation. In short, it is anthropomorphism—a naive, prescientific, and misguided attempt to understand animals, just as it was earlier claimed to be a naive, prescientific attempt to understand nature or God.

Given its rhetorical history, it is reasonable to begin to suspect that as a concept, anthropomorphism functions more as a rhetorical device than a logical analysis. And given this history, it is easy to understand many of the rhetorical features of anthropomorphism: It is a term of negative criticism that associates a type of understanding with an invalid or fallacious mode of thought, long since hounded from respectable intellectual company.

[There are some ethologists, Gordon Burghardt, for example, who have recently embraced as their methodology some version of what they call anthropomorphism. Burghardt (1990) argues for an anthropomorphic stance that he calls "critical anthropomorphism." What such definitions of anthropomorphism amount to is an attempt to exclude the negative criticism so prominent in ordinary uses of anthropomorphism by means of stipulation. In short, the meanings of the terms "anthropomorphism" and "anthropomorphic" are changed in this usage. This may not be wise given the negative connotations of these terms as ordinarily used. What these ethologists have in mind is to resurrect to limited respectability the kind of reasoning or inference that anthropomorphism refers to. Burghardt's notion, however, makes little attempt to defend the *truth value* of anthropomorphic explanations. Rather, he is interested in the *heuristic* value of anthropomorphic thinking in stimulating the imagination of the trained ethologist. So even if his proposal is sound, it is not a defense of anthropomorphism as ordinarily conceived.]

This history does not shed light on all of the rhetorical features of the concept of anthropomorphism. In particular, the association, very commonly charged by critics, between anthropomorphic thinking (or more generally, sympathetic treatment of animals) and appeals to emotions has not been accounted for. For example, consider the kind of suggestive rhetoric Spinoza uses when he says, "It is plain that the law against slaughtering of animals is founded on vain superstition and womanish pity than on sound reason" (*Ethics*, Pt. 4, prop. 37, no. 1. Quoted by Midgley, 1978, p. 352). The contrast between reason and emotion, which is such a central feature of the rhetoric surrounding the debate over animal rights and animal-minds is only partly accounted for by the points we have so far examined. I take it that the connection frequently made between

animal rights views and emotions is part of the same rhetoric that conditions the charge of anthropomorphic thinking. In a review of objections to primate research made by the animal rights movement a group of scientists conclude, "This illustrates how well-intentioned legislation based on emotional rather than empirical arguments can negatively impact scientific inquiry" (King, *et al.*, 1988, p. 1481). This denunciation of emotion and association of it with those who question such research is usually made automatically; in this case it is made in the same paragraph as the admission that "The campaign against primate research is actually based on the scientists' rationale for studying primates" (p. 1481). A similar accusation of emotionality and irrationality surely is carried by the charge of anthropomorphism.

IV. RHETORICAL COMPARISONS: THE PATHETIC FALLACY AND PERSONIFICATION

So far we have seen that the concept of anthropomorphism carries a double charge. In ordinary use it is treated partly like a fallacy, partly like a heresy. [Vicky Hearne (1987) anecdotally documents that the opposition to the way trainers talk about the animals they work with "suggests that modern injunctions against anthropomorphism have as much of a heretic-hounding impulse behind them as any of the older ones . . . for example . . . one person in the audience said that what I was saying sounded a little, well, religious. I patiently worked at finding out what she meant by religious, and it turned out that she meant 'anthropomorphic'" (p. 10). And later: "Academic psychology's role in this situation was, in the past, played by the Church. . . The nature of the offense of anthropomorphism—heresy—has not changed, only the institutions that perform the excommunications" (p. 265).] Like terms for both fallacies and heresies, "anthropomorphic" automatically carries a complicated and negative connotation. The rhetorical effect of the charge of anthropomorphism is to accuse some thinker of committing a mistake or fallacy that is well-known and exposed in the past, based on unsophisticated, uneducated thinking, and perhaps caused by emotions. The charge not only convicts someone of a mistake in thinking or reasoning, but it also suggests in an undefined way an explanation of this mistake—the cause or reason or motivation for making this mistake. So anthropomorphism is both a mistake and a mechanism, and this mechanism is associated, whether as cause or effect is not clear, with irrational emotion.

Looking at this history helps to characterize the rhetorical meaning of the charge of anthropomorphism. But it does not fully explain why it has this meaning. Nothing said so far provides a sound argument that "anthropomorphism" is merely a vacuous rhetorical weapon; the rhetorical effect

could well be based on a logical reality. It will be urged, indeed, that the explanation for the way “anthropomorphism” functions is to be found, simply enough, in the fact that it refers to a blatant logical mistake: the category mistake of ascribing human attributes to something not human. Category mistakes are made when one treats an entity of one type as if it were an entity of a different type, as in Gilbert Ryle’s (1949) famous example of seeing the buildings of the university and yet asking where the university is. Many assume that anthropomorphism is the category mistake of ascribing human attributes to something not human. (That this is how people implicitly think of anthropomorphism explains why it would sound odd to accuse parents of anthropomorphizing their infants).

This leads to the critical question: Does the logical reality of anthropomorphism match the rhetorical effect? Stated more precisely: Can we characterize anthropomorphism analytically in such a way as to justify the rhetorical implications we have discovered? For to ask whether “anthropomorphism” is a valid charge in the animal context is to ask whether there is in fact an intellectual mistake, reasonably thought of as anthropomorphism, which matches this rhetoric.

Let us ask this question concerning the simple definition of anthropomorphism broached a paragraph ago. Certainly it is an analysis of anthropomorphism; that is, it claims to tell us what anthropomorphism is and it justifies the rhetorical uses of the concept. If anthropomorphism is a blatant category mistake, then it is natural that it would be a criticism made casually and with the negative meanings we have explored. A criterion of an adequate analysis of anthropomorphism, however, is that it appropriately characterize those cases where the charge of anthropomorphism is invoked; that is, cases of ascribing mental states to animals. This the category-mistake analysis fails to do.

To show this, let us compare the charge of anthropomorphism with more obvious category mistakes, in particular with two familiar literary concepts: the “pathetic fallacy” and “personification.”

The *Dictionary of World Literary Terms* (Shipley, 1970) defines the pathetic fallacy (due to Ruskin) as the “presentation of the inanimate world as having feeling.” Standard examples range from “The one red leaf, the last of its clan/That dances as often as dance it can” to Pope, who sees all things as “. . . one stupendous whole, Whose body Nature is, and God the Soul” (p. 233). Clearly, the pathetic fallacy is closely linked to anthropomorphic and animistic capacities of thought. And like anthropomorphism, the pathetic fallacy is an unclear concept. The *World Dictionary* points out that “The usual meaning of the two terms that Ruskin wrings to special use make many suppose he is condemning where he but characterizes” (p. 234). The problem is that it is not clear whether there is anything wrong in such comparisons, and if there is, what it is. As one writer declares: “Pathetic fallacy reveals itself to be poetic truth”

(Shipley, 1970, p. 234). In a poetic context, in which expressions are not to be taken literally, it seems difficult to specify *in general* even for the purposes of poetic criticism what is defective about such comparisons. Of course, some might be more effective or more “truthful” than others, but that does not convict the class of comparisons of being fundamentally mistaken.

Let us ask not what is wrong with such comparisons but rather what are the criteria for labeling such a comparison the “pathetic fallacy.” It is clear that the term is used in cases in which the object being treated as human is of such a different kind (a leaf, a rock, the earth) as to make it obvious that the object cannot literally have the characteristics attributed to it. The difference in kind between humans and (say) leaves grounds the charge (carried by the terminology “pathetic fallacy”) that applying terms for human feeling to leaves is an obvious mistake if taken literally. But that depends on the sort of difference there is between the two types of entities being compared.

If it is correct to say that species have natures (and I will grant it for the sake of argument), then human beings have a unique nature, perhaps equatable with the human genetic material each of us carries. Suppose further that humans turn out to be the only creatures we know with various capacities, such as full-blown language and a certain sort of self-consciousness. Does this then settle the question? Is it then a fallacy to view animals as having human characteristics? This obviously depends on what we mean by “human characteristics.” If we mean those characteristics that define the unique human nature, then the answer is, yes. But which are those characteristics? If these characteristics are on the molecular level, then they have little to do with the sorts of comparisons in question between animals and humans. A biological answer to the question of how to define human nature, therefore, can hardly settle that issue. Surely it is an empirical question whether other animals, for instance, in fact have language. We have to look and see, and the same goes for the related question of whether they have the *capacity* for language.

The idea of the “pathetic fallacy” is that a category mistake has been made. But just because humans are in a different category than other animals, it does not follow that to compare them with other animals is a category mistake. Quite the contrary. We share many features, physical, biological, and social, with other animals, and it remains an empirical question which, if any, mental characteristics humans have uniquely. And even if scientists conclude that there are some, that other animals only have (say) signaling systems, while human languages are much richer (in syntax, for example), they had to look and see; it is not an a priori conclusion they can make in any other way.

Many philosophers (e.g., Descartes, Jonathan Bennett) have written as if the main issue concerning animal mentality is whether humans are unique in the animal world. [Bennett (1988) writes: “I have a real topic; there are differences of kind, and it is a real question which if any of them obtain between humans and

other animals" (p. 197).] And it seems a short step from arguing that humans have some unique property (or properties) to the conclusion that to think of animals as similar to humans is anthropomorphism. To put the position most tendentiously: if there is a difference in kind between persons (i.e., humans) and animals, is it not anthropomorphism to treat animals as if they were persons?

That way of putting it brings out the intuition that the charge of anthropomorphism rises or falls on there being a distinction between humans and animals. We can formulate it this way:

(*) There is a difference in kind between humans and other animals *if and only if* anthropomorphism (i.e., thinking of animals as similar to humans) is a mistake.

But, as we have just seen, anthropomorphism and a difference in kind between humans and other animals are logically separate concepts. What is confusing is the Cartesian heritage concerning (*). Descartes thought he had established a special distinction that attributed to humans all real mentality and to animals (i.e., machines) none. So, if we understand anthropomorphism to be the attribution to animals of mental terms (thoughts, feelings, intentions), then under that understanding, and if the distinction mentioned in (*) is the *Cartesian* distinction, then (*) would be true.

To a Cartesian, the comparison between animals and humans might seem like comparing humans and a leaf, and thus to be a category mistake, at least with respect to comparisons concerning an inner mental life. But this Cartesian dogma has long been discarded by scientists in favor of an evolutionary story of biological continuity. In the contemporary context, therefore, it would seem that it is at least a genuine question as to what attributes, say, primates possess. And if so, it cannot be a mere category mistake to attribute the attributes of human mentality to them. For it is not obvious that the attributes cannot literally apply to animals.

A use of language closely related to the pathetic fallacy can be called personification. (There is a different kind of personification associated with anthropomorphism that I will come back to in Section VII.) Here is an example of the type of personification I mean: Ingrid Newkirk (1988), in a forum on animal rights, says the following: "nature is cruel, but man is crueler yet" (p. 52). To describe nature as "cruel" could be considered a certain kind of anthropomorphism. And indeed the same is true of her use of "man." Both are better termed "personifications": treating an entity that is not a person as a person. *Man* is not a person, it is the set of human beings; both man and nature are abstractions. This leaves us with two questions: (1) Is this a mistake? (2) How does this relate to the charge of anthropomorphism in our relations to animals?

The answer to (1) is, not necessarily. Of course, if taken *literally*, it is a mistake. But it might be thought to be useful, if sloppy and dangerous, shorthand. No one literally thinks that man is cruel, but we do literally believe humans

are cruel, and this is a way of saying that. So it is not mistaken to assert that "man is cruel," but it is a mistake to believe it literally.

The answer to (2) is that it is easy to draw a sharp distinction between this sort of personification and the application of human terms to animals, though some may think of anthropomorphism as the same in both cases: applying person terms to nonpersons, that is, treating nonpersons as if they were persons. We can state it so generally as to encompass both: treating *something* that is not a person as a person. But in the case of "nature" it is so clear that this entity is an abstraction that utterances using it cannot be taken literally. When we look at what is being treated as a person we can draw a distinction. In one case it is an abstraction, in the other case it is a living nonhuman animal. It would be an obvious category mistake to take the first case literally. The second case is entirely different: we have a living subject which literally shares many properties with human persons, and at worse we are misapplying analogical reasoning. In the first case the mistake is clear and different from the alleged mistake in the second case. It could be argued that *man* and *nature* do not even literally exist.

These two cases thus differ in several ways. (1) The animal case surely involves two entities of the same biological and physical type. (2) The animal case clearly involves an *inference* from the animal's behavior and appearance to the application of human terms. No such inference is being made in the abstraction case. A literal comparison between nature and humans cannot be made to begin with. (3) In the animal case it seems that what we are doing is postulating mental states as causal entities to explain the animal's behavior. This is not paralleled in the case of personified abstract entities. "Nature is cruel" does not explain.

In both examples, pathetic fallacy and personification, real category mistakes would be made were they to be taken literally. But the category-mistake model of anthropomorphism does not appear to apply to the cases in which the charge is commonly made, since clearly they differ in fundamental ways from real category mistakes such as would be exemplified by literal readings of personification and examples of the pathetic fallacy. So here the rhetoric matched the analysis, but the analysis is not plausible.

V. A LOGICAL FRAMEWORK FOR ANALYZING ANTHROPOMORPHISM

There are three questions we need to answer to give a logical characterization of anthropomorphism:

1. What is it? That is, what sort of mistake is it?
2. When does it occur? Under what conditions do we make this mistake?
3. Why is it a mistake?

Different answers to these questions yield different conceptions of anthropomorphism. I believe that thinkers do often have different conceptions of anthropomorphism without being aware of that fact. To resolve the issue of whether there is a match of rhetoric with logic, we not only need to provide a clear and consistent set of answers to these three questions, but our answers must be such as to justify the negative rhetoric that informs the charge of anthropomorphism.

With respect to what anthropomorphism is, I propose that there are two broad categories: *interpretive anthropomorphism* and *imaginative anthropomorphism*. Before explaining these it will be useful to define the class of extended mentalistic predicates: predicates referring to mental states and processes, cognitive and emotional, as well as verbs of action (e.g., “hunt,” “play”) and predicates of moral character and personality (loyal, brave, sneaky). Call these M-predicates.

I intend that interpretive anthropomorphism refer to all of the usual cases of ascribing M-predicates to animals on the basis of their behavior. Interpretive anthropomorphism, then, refers to cases of inference, from animal behavior to M-predicates, where these include descriptions of the animal’s physical behavior in terms of intentional actions (the deer are playing, the chimp is trying to solve the puzzle). So the sort of thing anthropomorphism is in these cases is an *inference* from the animal’s behavior described in terms of bodily movement; it also could be considered an *interpretation* of the animal’s behavior or an *explanation* of the animal’s behavior based on positing the truth of the ascribed M-predicates. The variations do not matter since the inference (or interpretation) is meant to be a sort of explanation. Since the charge of anthropomorphism always assumes that a mistake has been made, we can say that anthropomorphism on this conception is a fallacious inference. Interpretive anthropomorphism is surely the intended target of criticism by philosophers and scientists when they claim that anthropomorphism is a mistake to be avoided in explanations of animal behavior. Like any other fallacy, we need an account of the conditions under which it is committed and an account of why it is a fallacy.

By imaginative anthropomorphism I mean the productive activity of representing imaginary or fictional animals as similar to us. Examples of such representations are the animal characters in animations, books, drawings, movies, and stories that are treated in the story as similar to us, not only in having M-predicates, but in behaving in a similar way, for example, speaking, and so on. This form of thought is obviously intertwined with interpretive anthropomorphism, and while not directly targeted by philosophers and scientists concerned about scientific explanations of animals, it can just as properly be called anthropomorphic. It, too, sometimes elicits complaints from critics. In a review of John Crompton’s novels about social insects a critic notes of Crompton’s tendency to make insects talk, “Superficially, this smacks of anthropomorphism, anathema to today’s nature writers (Kendrick, 1988).

Humans obviously have an enormous ability to treat representations as personlike, whether of raisins, trees, or little engines that could; all that is required is to paint on a human face. It is plausible that some of the disrepute and suspicion of interpretive anthropomorphism comes from our ability and desire to produce such representations.

Before returning to interpretive anthropomorphism, I want to consider the suggestion that there is a third main type of anthropomorphism. It might be suggested that in addition to anthropomorphic inferences and representations there are anthropomorphic beliefs; that is, a set of beliefs about animals that amount to anthropomorphism. I want to resist this suggestion. It is true that there are beliefs, or propositions held, about animals that link animals and M-predicates, and that these propositions are denied by tough-minded thinkers—for example, that bees have a language, that dogs have beliefs, that chimps have self-awareness. But anthropomorphism is a particular kind of mistake, it is not just false belief. And the mistake, if it is one, has to be one based on misinterpreting animal behavior. As Jonathan Bennett (1988) says, “Descartes must be granted this much. We attribute mentality to others on the basis of their behavior, and no other basis looks even remotely plausible” (p. 199). Bennett in fact leaves out the important biological evidence of physical constitution and evolutionary history, without which I do not think we would even know how to describe the behavior. But the point still stands that we always make an inference in ascribing M-predicates, and the inferences that elicit charges of anthropomorphism are always based on the animal’s behavior. Vicki Hearne’s (1987) defense of “anthropomorphic” descriptions of animals is clearly based on behavior, in particular on how animals respond to us: “trees don’t answer in the way animals do, and most of animate creation doesn’t answer as loyally and with as much respect for the details of the human landscape as dogs, cats and horses do” (p. 264). Ultimately, if some beliefs, such as those mentioned, are to be considered anthropomorphic, it must be because they have been mistakenly inferred from the behavior of such animals. If behavior is all we have to go on, then such beliefs cannot in themselves be absurd, or known a priori to be false. They must be defective as explanations of the animal’s behavior.

With a view toward clarifying the way philosophers have talked about anthropomorphism, I propose making a further distinction, subdividing interpretive anthropomorphism into *categorical* and *situational* anthropomorphism. Categorical anthropomorphism is applying M-predicates to creatures to which they do not apply under any of the behavioral circumstances in which the creature is ever situated. Categorical anthropomorphism is almost like making a category mistake; relative to the type of creature, it is always a mistake to ascribe a particular M-predicate, a mistake of interpretation. By contrast, situational anthropomorphism happens when we, as we sometimes do, misinterpret an animal’s behavior in ways that could correctly apply to that animal, but which do not apply in the situation in question. An observer might, for example, interpret a

chimp's show of teeth as a sign of anger when it is (let us imagine) a sign of affection. Situational anthropomorphism is relative to the situation; the M-predicate chosen, while mistaken, is not categorically inapplicable to the animal in question. Categorical anthropomorphism, on the other hand, occurs when one ascribes on the basis of behavior an M-predicate that never applies to this sort of creature. The concept of categorical anthropomorphism articulates the idea that there are inherently misguided ways of taking about animals, inherently misguided expressions to use in describing animals.

We can further subdivide categorical anthropomorphism in terms of the conditions under which it is committed. (1) *Species type*: application of mentalistic predicates could be counted as anthropomorphism depending on the species of animal. What wouldn't be anthropomorphism concerning a chimp might be concerning a worm. (2) *Predicate type*: application of mentalistic predicates could be counted as anthropomorphic depending on the predicate. I have in mind applying the wrong types of predicates, out of the extended set of mentalistic predicates, for example, predicates of moral character, to a given creature. While perhaps it can be said that a horse is trying hard to win the race, perhaps it cannot be said that he is courageous.

So, to answer whether any use of mentalistic predicates is anthropomorphic we need to consider both the creature and the predicate applied. A clear example of a claim involving both predicate and species comes from Peter Carruthers (1989) who claims, "only the most anthropomorphic of us would be prepared to ascribe second-order beliefs to toads and mice; and many of us would have serious doubts about ascribing such states even to higher mammals such as chimpanzees" (p. 261). Clearly there is a range of potential targets for the charge of categorical anthropomorphism, as well as a range of positions available to critics of anthropomorphism. The most extreme critical position seems also to be the common one: I will call (on analogy with hard determinism) *hard anthropocentrist* (*hard centrists* for short) those critics of anthropomorphism who regard the application of *any* M-predicate (perhaps with the exception of pain and other sensations) to *any* nonhuman animal as categorical anthropomorphism. [Carruthers (1989) even denies that other animals have experiences.] Hard anthropocentrists see categorical anthropomorphism as a universal problem. Hard anthropocentrists are committed to a sharp division between humans and other animals in which, perhaps excepting pain and other sensations, it is always a mistake of categorical anthropomorphism to attribute M-predicates to animals and never a categorical mistake to ascribe them to humans. Hard anthropocentrism seems to be the standard position among tough-minded philosophers and scientists, certainly among those who freely make the charge of anthropomorphism.

It is important to keep in mind that I have defined anthropocentrism relative to the charge of anthropomorphism. Anthropocentrists are not just those who

believe as a matter of empirical fact that certain M-predicates do not truly apply to certain animals. In addition they must hold that to ascribe the relevant M-predicates, on the basis of the animal's typical behavior, is to commit a mistake of categorical anthropomorphism. Donald Davidson, for example, is anthropocentric in my sense *only* when he suggests that those who are inclined to ascribe thoughts to animals on the basis of their purposive behavior are anthropomorphic (not just wrong). To claim, for instance, that it is anthropomorphic to ascribe self-consciousness to a chimp is, presumably, to say more than merely that the ascription is false. For otherwise, the disagreement might just be empirical, one about which conscientious investigators could disagree, and about which they could change their minds. Anthropomorphism, with its rhetorical implications, need not enter in. Consider, for example, those who believed on the basis of his performances, and before the mechanism of cue transmission was discovered, that Clever Hans could add numbers. Such observers were mistaken about Hans' mathematical comprehension, but there was not anything necessarily anthropomorphic about their reasoning.

Some of us are less extreme than hard anthropocentrists, but we too would hesitate to apply mentalist terms in some cases, for example, to individual insects or to the hive in the case of social insects. So most of us think it *can* be categorical anthropomorphism to apply most M-predicates to some species, for example, worker ants. Also there are some M-predicates that we would hesitate to apply to any nonhuman animals [e.g., Wittgenstein's (1953, p. 174) famous example of a dog believing that his master will come the day after tomorrow.] Thus there are a whole range of positions open to the anthropocentrist concerning where to draw the line, relative to species and predicate, in making the charge of anthropomorphism. Call these positions less hard than hard anthropocentrism, *modest anthropocentrism*. Most of us are at least modest centrists about some species and certainly about some predicates; most of us think that it is possible to make the mistake of categorical anthropomorphism in some cases, for example, attributing intentions to an army ant hive.

VI. SITUATIONAL ANTHROPOMORPHISM

Situational anthropomorphism as I have defined it is a misinterpretation of an animal's mentality based on the animal's behavior and some sort of analogy to human behavior. In particular, it is the ascription of an M-predicate that *could* be true of the animal, but in fact is not. In short, a factual mistake is made. As long as the basis for the inference is in part a perceived similarity between the animal's behavior and human behavior, this is rightly considered a sort of anthropomorphism.

The description I have given of situational anthropomorphism agrees with Mary Midgley's (1978) description of anthropomorphism per se:

Now people obviously sometimes do attribute to animals feelings and motives that they do not have, treating animals as if they were people. And this is called *anthropomorphism*. The ordinary use of the term, however, implies that the mistake is one of fact; there are other feelings the animal *does* have, and we can often name them. Thus the sentimental dog owner attributes affection to his pet when bystanders attribute greed. (p. 344)

Midgley's description could be taken to refer to categorical anthropomorphism (of the predicate type). I do not think she means to say that affection cannot be attributed to any dog, however, but rather that *this dog* is not truly affectionate, or at the least this example of the dog's behavior is not motivated by affection.

There are several different kinds of cases of situational anthropomorphism. Before spelling out the distinctions I have in mind, it will be useful to characterize the common feature of all interpretive anthropomorphism, both situational and categorical: the form of explanation. We ascribe M-predicates to nonhuman animals by way of an explanation. We explain animal behavior as we would explain our behavior if we (humans) were in *similar circumstances*. (Where circumstances can include: sensory stimulation, behavior, bodily attitude, history, social setting. Of course, it is not easy to define "similar circumstances.") I will call this interpretation the "anthropomorphic inference." Clearly the analogy between humans and animals must be quite elastic. Consider elephants. Eltringham reports that he has "often seen a mother fall back and offer a helping trunk to enable it to scramble up a steep bank" (1982, p. 82f). Here we clearly feel an analogy between the trunk and a human arm, so the anthropomorphic inference is based on taking the elephant's trunk and human arms equally to be appendages directed to helping.

In all cases of situational anthropomorphism we are misinterpreting animal behavior when we treat it as understandable by analogy with our behavior. The intuitive characterization of situational anthropomorphism is that it is like the sort of mistake a naive observer can make about an alien culture. A head shake that in her culture means "no" means "yes" in the alien culture. Situational anthropomorphism is thus a kind of anthropological misinterpretation. In this regard Asquith's (1986) contrast between Japanese and Western primatologists is illuminating: "The Japanese approach primate studies from a cultural rather than a biological point of view" (p. 65). The Japanese would be concerned to avoid errors of situational anthropomorphism whereas Western animal scientists are focused on avoiding categorical anthropomorphism.

Case (1). The first distinction to make is between misinterpreting a unique particular bit of behavior and misinterpreting a type of behavior. Most scientific interest in animal behavior focuses only on types of behavior. Indeed, while most of us look at our own behavior as a series of unique actions having a dramatic

structure and a unique context that explains what we did and why we did it, we see animal behavior as a series of types of behavior from a relatively small set of types. But there is no obvious reason for denying that animals can act in ways that merit consideration as unique acts, just as we regard our own actions. So, it is theoretically possible to misinterpret a particular bit of behavior and ascribe an M-predicate that could apply (relative to species and predicate) but does not. This yields the first type of situational anthropomorphism.

Let us look at an example. Vicki Hearne (1987) tells the story of a police tracking dog, Rinnie, who was dispatched to track down the kidnapper of a supermarket worker: "After taking a good sniff, the dog calmly and without hesitation, walked around the car to where the victim was talking to the police officers, and bit him in the seat." Rinnie's handler was "chagrined . . . and the dog was disgraced" (p. 26). As the story turned out, the worker had robbed the supermarket and faked the kidnapping, and the dog was correct after all. Rinnie's behavior struck observers at the time as a sign of some sort of instability or aggression. We could say, then, that this particular bit of behavior, which Hearne obviously takes to show that the dog somehow understood that the victim was responsible, was misinterpreted as (say) a feeling of hostility. This is case (1) situational anthropomorphism if the interpretation is based on an analogy with human behavior. But this is not clear since we are not dogs and we rarely bite other people. So it is not clear here whether the interpretation is based on an anthropomorphic inference.

Take a clearer example of this subcategory of situational anthropomorphism. In the movie, *Lassie Come Home*, Lassie swims across a large river and drags herself up on the shore exhausted to return to her master. In reality we learn that the dog that played the part was trained to behave as if he were exhausted (in fact, exactly as a human would behave if he crawled out of a river on his hands and knees) to a series of hand commands from his handler. I take it that our tendency to misperceive or to perceptually misinterpret the behavior of Lassie is one sort of example of anthropomorphism played upon by animal trainers in the entertainment business.

To look at the Lassie example as a general lesson about anthropomorphism would be mistaken. It says nothing about whether a dog could swim across a river and drag himself up on the shore exhausted, that is, feeling exhausted. It's true as well that the dog playing Lassie only "acts"—it is a fiction and so the dog is not "really" doing *any* of the things we see on the screen. But what does this really prove? That dogs do not, in fact, do any of those things? It only shows that we can be fooled. Nonetheless, tough-minded thinkers would like to suggest that animals behave in their various instinctive ways that we for some reason—due to our anthropomorphic programming—systematically misinterpret along human lines. Just as the dog crawling along the ground is not "exhausted," so dogs and

other animals behaving in ways similar to humans are not really the other traits or do not really have the mental states we anthropomorphically attribute to them in real life.

Case (2). Focusing now on types of behavior, a second type of case would be to anthropomorphically mistake one type of behavior for another, perhaps equally intentional, conscious, or whatever. It is not that no M-predicate applies, but rather that we attribute the wrong M-predicate, and do so because of an anthropomorphic interpretation.

Consider examples from the elephants. Trunk greetings could be signs of affection or attempts to gain information. Eltringham (1982) says, "In greeting each other . . . the trunks are extended and the tips touch rather like two people shaking hands, except that more information, in the form of scents, is exchanged than is the case with a handshake" (p. 65). But perhaps elephants do not have our concept of a greeting, and the correct understanding has more to do with the exchange of information. So there are at least other possible interpretations.

Or consider what are called "burial gestures." Cynthia Moss (1988) tells us that when an elephant dies the other members of the family attempt to cover up the corpse:

They stood around Tina's carcass, touching it gently with their trunks and feet. Because it was rocky and the ground was wet, there was no loose dirt, but they tried to dig into it with their feet and trunks and when they managed to get a little earth up they sprinkled it over the body. . . . By nightfall they had nearly buried her with branches and earth. Then they stood vigil over her for most of the night and only as dawn was approaching did they reluctantly begin to walk away, heading back toward the safety of the park. (p. 73)

It is difficult to resist concluding that the elephants mourn their dead and are sad about it (especially given the way it is described: "stand vigil," "sprinkled dirt," "reluctantly begin"). But perhaps, although they can be said to be sad in other circumstances, it may be the case that they fear the dead and are trying to get rid of her. I do not suggest that this is a better or even very good explanation, but only that it might *possibly* be a better one. Further observation might determine which is the better hypothesis.

Case (3). In a different category, and more difficult to imagine, would be cases where we anthropomorphically ascribe M-predicates when the true explanation is really very different from human mentality. Here I am thinking, not of reductionist answers, that argue away the animal's mentality but rather of the possibility that the animal's mentality is very different from our own, though in some sense on the same level. Moss (1988) suggests something like this when after describing the exuberant behavior of elephants when they meet family members she says:

After 18 years of watching elephants I still feel a tremendous thrill at witnessing a greeting ceremony . . . I have no doubt even in my most scientifically rigorous moments that the elephants are experiencing joy when they find each other again. It may not be similar to

human joy or even comparable, but it is elephantine joy and it plays a very important part in their whole social system. (p. 125)

Moss suggests that we be anthropomorphic and yet allow for the real limitations of our anthropomorphism; the suggestion is that elephant behavior does indicate joy, but a joy very different from our own. I think sense can be made of this, but it implies that an ordinary anthropomorphic response to elephant greetings would be at least partially mistaken.

[An example that baffles our anthropomorphic intuitions while making clear the possibility of, as it were, alien M-predicates concerns the way elephants respond to elephant skeletons. Moss (1988) says "They seem particularly interested in the head and tusks. They run their trunk tips along the tusks and lower jaw and feel in all the crevices and hollows in the skull. I would guess they are trying to recognize the individual" (p. 27). But she also tells of "Agatha of the AA [family who] stops and feels and gently moves her mother's skull, even though it is several years after Annabelle's death and Agatha passes the spot frequently" (photo caption, n.p.). It seems that something very meaningful is going on for the elephants and it is on the M-predicate level if anything is. But it is so alien to us that anthropomorphic intuition is rebuffed.]

Case (4). To explain the next category, I need to first introduce the concept of a P-predicate. P-predicates ascribe the kinds of properties favored by behaviorists and others inclined to nonmentalistic sorts of explanations of behavior. A P-predicate will be a predicate relating to a physical level of description; examples would be bodily processes (including sensory responses) described physiologically, operant conditioning schedules, and so forth. Assuming for the sake of the example that animal training operates on the nonmentalistic level, the training in the Lassie example counts as a P-level explanation of what is really going on. Lassie does not feel exhausted; Lassie is responding to his handler's cues.

Even common sense explains some human behaviors in terms of P-predicates. We have automatic reflexes and bodily processes that are not intentional, and so not to be explained by M-predicates. This generates another possible type of situational anthropomorphism, type (4): It is possible to miss an example of animal behavior that calls for explanation by M-predicates because we anthropomorphically assume that it is to be explained by P-predicates, again projecting human experience too strongly onto the potentially alien experiential world of the animal. A possible example is the elephant tummy rumble. Some observers think the tummy rumble is a form of intentional call. Eltringham (1982) thinks not: "Elephants certainly have noisy digestive systems but it is doubtful that intestinal churning has any direct function in communication" (p. 66). Perhaps so, but again, this could be a prejudice in favor of sounds produced with the vocal chords, a kind of reverse anthropomorphism. Since our tummy rumbles do not mean anything, the elephants' do not either.

Types (1)–(4) are conditions under which situational anthropomorphism could occur. Before indicating a final category, which shades into categorical anthropomorphism, it will be useful to ask why situational anthropomorphism is a mistake. I have sketched examples where mistake of the types specified are *possible*. Hence, at least in these sorts of cases, mistakes are a danger. Were the anthropomorphic inference to be the final judgment, we would or could be mistaken because there is an *alternative explanation*. In the Lassie case a P-predicate explanation is correct, in the Rinnie case, perhaps a different M-predicate explanation is correct. This is true in the other cases as well. There are alternative explanations, difficult to formulate perhaps, that might be better or more nearly correct. In several of these cases it might be difficult to even say what would count as investigating the alternative explanation or gaining evidence of its truth. But on the assumption that we can indicate what would count, we can see what would be dangerous about resting with the anthropomorphic explanation. Still, nothing said so far implies that in cases of potential situational anthropomorphism a mistake is *necessarily* made, that to make the anthropomorphic inference is in itself a blunder. We have only seen how it is *possible* that such an inference goes wrong.

Does the logic of situational anthropomorphism match the rhetoric? Surely not. Mistakes can be made this way, but as we saw earlier, anthropomorphism is supposed to be a fundamental and obvious mistake. It is not just proposing the wrong explanation, but proposing the wrong *type* of explanation. Yet nothing said so far implies there is in principle anything wrong with the proposed explanations involved in situational anthropomorphism; they may be wrong in particular cases, as might any other hypothesis. Therefore more observations and tests may be called for to be responsible. But equally the charge of anthropomorphism in these situations needs to be backed up by suggested alternative hypotheses to be a responsible charge.

As other writers have pointed out, the critic of anthropomorphic thinking wants to generalize particular mistakes into a universal mistake. Midgley (1978), for instance, alludes to this generalizing impulse in her discussion of anthropomorphism: “It is a much more extreme position to say that we would be wrong, not just in attributing affection to *this* dog, but in attributing to any dog any feeling or motive whatever” (p. 344). Only if such a generalization can be made will we have a mistake worthy of the rhetoric that infuses the meaning of the charge of anthropomorphism. How is this generalization to be made?

The most typical way is to claim that the M-predicates in the anthropomorphic explanation *could* not apply. But this is not situational anthropomorphism. The difference is easier to see if we look at the final type of situational anthropomorphism:

Case (5). This type of situational anthropomorphism occurs when we anthropomorphically ascribe an M-predicate explanation when a P-predicate one is

in order. A potential example of this type comes from Eltringham’s (1982) explanation of the coordination of elephant sleeping: “No one individual appears to decide when bed-time has arrived but all lie down or stand still within a few minutes, even if they appear to be out of sight of one another. I could not resist the conclusion that, in some way that I could not detect, they were able to communicate their intentions to each other” (p. 11). One can imagine alternative explanations to the one Eltringham favors: perhaps the elephants somehow coordinate their eating so that they all feel sleepy at the same time; perhaps they emit some sort of pheromone that causes them to feel sleepy; and so forth. Surely there are alternative explanations, and probably explanations at a lower P-level, to the idea that the elephants engage in intentional communication in this case. But none of this is to say that the analogy that plays a role in Eltringham’s thinking is necessarily wrong, nor is it to say that there are not other cases in which elephants engage in intentional communication. The M-predicate, “communicate intentions,” could apply to elephants, I assume; perhaps it does not in this case. And this is what all the cases of situational anthropomorphism, (1)–(5), have in common. They are all mistakes about particular cases of application of M-predicates. They are potential misapplications of an M-predicate whose application to these animals is not wrong in principle. But many thinkers would resist the elephant case and many of the other cases of attributions on the grounds that they are wrong in principle, that the M-predicates *could not* apply to the animals in question. And that is categorical anthropomorphism.

Does the existence of situational anthropomorphism imply the validity of categorical anthropomorphism? Centrists might think so. Their thought is that under the influence of anthropomorphic thinking we suffer from an illusion; we always mistake animal behavior as we do in the Lassie case, or as perhaps Eltringham is doing in taking a physical process for communication of intentions. Consider sled dogs. They seem extremely “loyal” and “devoted” to their trainers and will make incredible efforts for them; these are the M-predicates we are inclined to apply. Yet, it seems that sled dogs have a strong innate desire to pull and to run in a pack. Perhaps all the trainers do is to shape their behavior to the behavior they need to get from the dogs in order to run races. The trainer is the alpha leader and the dogs are the pack who behave instinctively. The generalization is that an animal’s behavior, which has a functional and/or evolutionary (P-level) explanation, is systematically misinterpreted by humans in terms of commonsense explanations of human behavior.

Can the Lassie parallel be extended to the sled dogs? Is there an illusion like that created by Lassie? Are we misled in some systematic way by their pulling? For example, are the dogs not really loyal? Are they *just* behaving in ways that descend from their wolf pack behavior? To make a serious case here, one would have to look at what we mean by loyalty. And clearly what we mean by loyalty is systematically constituted by the very sort of behavior that these dogs exemplify:

pulling till they drop, defending their master, returning to their home base by themselves, affection (even if doglike in form), not doing what other people want (not obeying orders of a stranger), and so forth. Some dog behavior is even so paradigmatic of loyalty it is hard to believe that loyalty is solely a human-centered concept as opposed to one in which animal behavior has always constituted a paradigmatic example. The only way the tough-minded argument can work concerning loyalty is by assuming that there is *nothing* like a primitive version of our mentality lying behind the dog behavior. This amounts to viewing dogs as like automata that are programmed with blind instinctual behaviors that happen to be useful to us, and which we misperceive as flowing from a mentality similar to our own. Under this assumption we misinterpret the behaviors as exemplifying loyalty almost as we misinterpret Lassie's crawling up the bank as exhibiting feelings of exhaustion. Since this assumption is logically equivalent to categorical anthropomorphism, it cannot be used in a noncircular way to argue that the existence of particular mistakes (situational anthropomorphism) logically implies the existence of a general mistake (categorical anthropomorphism). Categorical anthropomorphism needs to be motivated independently.

VII. CATEGORICAL ANTHROPOMORPHISM

To hold that certain M-predicates could not *in principle* explain animal behavior or be properly ascribed to certain animals is to shift to categorical anthropomorphism. Categorical anthropomorphism occurs when, for a certain predicate and species, the anthropomorphic inference is mistaken in principle. Consider this example: When prairie dogs meet they greet each other with an apparent kiss. The anthropocentrist will argue: Surely it is a mistake in principle to think that prairie dogs really kiss, that they kiss as we kiss, with similar intentions, thoughts, emotions, and so forth—that what they do can literally be said to be a *kiss*. And this is not just a question of prairie dogs having other intentions (like an unfamiliar tribe) or of having strange or alien intentions (like Martians), but that there are simpler explanations (e.g., it is merely a sign of recognition) that do not attribute any of the rich complex of intentions, cognitions, feelings, or social relations associated with kissing. It is only anthropomorphism to think that superficial behavioral similarities are to be understood as caused by underlying mentality of a familiar sort.

Generalize the reasoning just formulated and you have the usual theme of philosophers who have worried about animal mentality. Most philosophers have tended to think there is a sharp division between humans and other animals—either humans have minds and animals do not, or humans are rational and animals are not, or humans use language and animals do not—such that it is a

mistake in principle to apply various M-predicates to animals. And always this division is a hierarchy, with humans at the top and animals underneath separated by a gulf. Thus philosophers have not been concerned with getting the animals' mentality right, that is, to avoid the anthropological error of mistaking the expression of one mental state for that of another; nor have they been concerned to ban anthropomorphic interpretations to make room for another sort of mentality, alien to our understanding. Rather, they have aimed to ban anthropomorphic interpretations in order to deny to animals any sort of mentality. Of course, each philosopher draws his line at a slightly different point. Davidson (1975) worries about anthropomorphism in connection with ascription of any thought to animals; others might draw the line more or less extremely. It would be fair to sum up this tradition, however, as one that supposes the dangers of anthropomorphism are quite general; that is, the charge of categorical anthropomorphism is made in the case of most M-predicates and most species. In short, when philosophers have been concerned about anthropomorphism, they have tended to adopt ways of speaking that make them sound like hard anthropocentrists.

Having already described the sort of mistake categorical anthropomorphism is supposed to be, I will focus here on the reasons why attribution of M-predicates to animals (i.e., the anthropomorphic inference) is thought to be the mistake of categorical anthropomorphism. I will examine not only the few discussions of anthropomorphism by philosophers but in addition other potential objections that I believe implicitly lie behind the categorical rejection of anthropomorphic inferences.

A. Anthropomorphic Inference Is Connected with Falsehood

It is safe to say that most references to anthropomorphism in the scientific and philosophical literature occur in discussions of whether animals think or have other M-states. As I have maintained in preceding sections, the question of whether animals think and the issue of anthropomorphism, though related, are logically distinguishable. Those who fail to see this tend to assert that it is anthropomorphism to ascribe M-states to animals. On the contrary, an ascription of M-states to animals is not necessarily anthropomorphism even if it is false.

The idea that anthropomorphism is simply false belief or assertion can be spelled out at two points in the anthropomorphic inference. Some may argue that ascription of M-predicates is anthropomorphism because it is based on false premises or assumptions. Others have the idea that it is anthropomorphism because it draws a false conclusion. I shall argue that neither of these is anthropomorphism in the proper sense.

Consider, first, the idea that the anthropomorphic inference is based on a

false assumption, an assumption of similarity between the animals in question and humans. The question is: Why is this a false similarity? Human physiology and human behavior have both evolved from that of other animals, so it is not obvious that the analogies that ground the anthropomorphic inference are not a valid basis for it. Indeed, it appears that in attacking the assumptions underlying the inference the centrist who charges anthropomorphism gives a circular argument. In his early book on animal behavior, *Rationality*, Jonathan Bennett (1964) suggests as much:

[I]t begs the question to say that it is merely "sentimental" or "anthropomorphic" to credit bees with rationality. If "It is anthropomorphic to credit bees with rationality" means "To credit bees with rationality is to liken them to humans," then this is true but unhelpful; on the other hand, if it means "To credit bees with rationality is wrongly to liken them to humans," then it begs the question. (p. 11)

The very point at issue is whether it is mistaken to draw the anthropomorphic inference, so an argument needs to be given by the centrist as to why the general analogy is mistaken, and mistaken, not just in this case, but in principle. It is difficult to see how such an argument would go. Whether or not the analogy is a good one seems to be an empirical question, and as such, the claim that it is not a good analogy, even if sustainable, would not ground a charge of anthropomorphism but merely one that the inference is mistaken.

Consider some particular assumptions that underlie anthropomorphic inferences. For example, we assume that the elephants who appear to greet each other know each other and have appropriate social relations. For such social relations would seem to be required for the application of our concept of a greeting. Whether such assumptions are true is an empirical question, however, and hence the assertion that they are not true does not justify the charge of categorical anthropomorphism, which holds that the anthropomorphic inference is wrong in principle.

Sometimes thinkers reason that if the conclusion of the anthropomorphic inference is false, even empirically false, then the inference is guilty of categorical anthropomorphism. This way of thinking equally misuses the charge of anthropomorphism to rule out what might be a reasonable empirical dispute. Consider the Clever Hans example again. It is conceivable that there was another and smarter horse called Clever Shmans; a horse that, as far as anyone could tell from his behavior, could add numbers and did not use cues from spectators. How could it be asserted that to accept the claim of Shmans' ability is anthropomorphic? It is possible that we might argue on empirical grounds that Shmans cannot add. For example, it might be claimed that there is something missing in a horse's brain that is required in the human brain to understand numbers. If such evidence were found, however, it would not support a charge of anthropomorphism against those who accepted the ascription of numerical ability on the basis of Shmans' behavior.

B. Anthropomorphic Inference Has Suspect Origins

One of the most interesting reasons for being a hard centrist can be located in the general role that anthropomorphic inferences play in human cognitive development. The connection of anthropomorphism in literal thinking about animals to imaginative anthropomorphism and of imaginative anthropomorphism to imagination and childhood provides a motive to question anthropomorphic inference and to see it as categorical anthropomorphism.

Children invest dolls and other objects with personalities as part of normal cognitive development. They see personalities everywhere. Even when we grow up we still easily see human forms in clouds, cracks, cars, and so on; indeed, much visual art is based on our ability to see human forms in vague lines and shapes. Clearly humans have a capacity to perceive objects as having personalities. It is not just a capacity to imagine; it is a sort of imaginative representation of a perceived object. As children we often accept these representations as real.

When this capacity is exercised on animal representations, I think we have imaginative anthropomorphism, which must be separated from our more general capacity to imagine personalities in things. Without this distinction the acceptance and enjoyment of nonhuman entities as fictional persons in itself would be anthropomorphism. But if someone enjoys Kermit the Frog or the Little Engine That Could, is that anthropomorphism?

One is inclined to say that just to *accept* these characters is not anthropomorphism, because to accept them requires no beliefs at all. Instead it may be like *treating them as* or *seeing them as* people. I need to be facilitated in that perception; they have to constantly look and act like people. But I don't have to *believe* that frogs are persons in order to understand the character of Kermit the frog any more than I have to believe that trains are persons in order to understand the Little Engine That Could.

To apply the concept of anthropomorphism to fiction, we can crudely divide fiction up into the realistic (R) and nonrealistic (non-R). Imaginative anthropomorphism could then be said to be involved in those fictions that have animal characters but are implicitly labeled R (or *not* labeled non-R). Or it could be said that anthropomorphism resides in the viewer and defined: a viewer is guilty of anthropomorphism when in reacting to a non-R animal fiction he or she

(*) Fails to be clearly aware that it is non-R

You know when you read the Little Engine That Could that there cannot be such a train; we are *pretending* that such a train is trying to climb the hill. But a very small child might not know that there *could not* be such a train. To understand and engage the story it is not necessary to know it is possible. It is only necessary to pretend.

The closer something gets to human form and behavior, the easier it becomes to pretend it is a person. But whereas we have to pretend that a train *could* be a person or personlike, the worry about Disney's *Bambi* might be that it encourages young viewers to think that it is merely an R fiction; they do not have to *pretend* that it is possible, they believe in an unreflective way that it is possible, just as I believe that the background events and characters in some novels are possible. Thus there are (very roughly) two levels of beliefs and pretense in our response to fiction:

1. That these particular individuals and events *exist*
2. That these individuals and events are *possible*

In *The Old Man and the Sea* by Ernest Hemingway, I don't pretend that the events are possible or that certain kinds of creatures exist. But I do so pretend when I read the *Little Engine That Could*. By supplying the requisite properties, the artists and writers make it easier to pretend, or if you will, present a world in which there is a talking train, and we pretend that that world exists. But a young reader might think that it was our real world. According to the suggestion in (*) that would be anthropomorphism.

To summarize these thoughts about representations and imaginative anthropomorphism, we can say that fictions are anthropomorphic when they represent animals as more like humans than in fact they are. But spectators are often not taken in by such representations, nor are they influenced by them. Even children know that raisins aren't really like the California Raisins and that trains are not alive. Spectators are guilty of anthropomorphism only when they accept that the false nearness to humans is possible. When dogs talk in Richard Adams' movie *The Plague Dogs*, this is anthropomorphic, but viewers are not taken in. In Disney's *Bambi* most of the other details of forest life, in contrast to *The Plague Dogs*, are entirely misrepresented. If viewers are taken in, and accept the fiction as realistic, they are guilty of anthropomorphism.

Our general capacity to understand such anthropomorphic representations makes suspect the way we understand real animal behavior. The hard centrist may wish to suggest that commonsense explanations, in terms of M-predicates, are too closely allied to childish anthropomorphism, indeed dependent on those imaginative capacities, that we all grew up with and which we regularly apply in imaginary fictions. Just as we can easily imagine a Bambi on screen as a pre-adolescent human child, so when we see a real deer there probably is a certain amount of imagination going on in how we relate to and understand the deer. It seems likely that imaginative representations play an important role in how we understand animals commonsensically. Surely these imaginative capacities are often misused; and just as surely they do not provide the sort of foundations we hope to have for scientific explanation.

That we are able and inclined to see personality in some animals, at least,

can even be given an evolutionary explanation. William Calvin (1983) suggests that human affection for cats has an evolutionary explanation in protective mimicry:

One has only to observe a human holding a cat to realize what is going on: the pet is invoking the same reactions that a cuddly baby sets off. Their contented responses when cradled set off the same flood of emotions in us. Babies babble and nuzzle, cats purr and rub. (p. 16)

The idea is that there "is a symbiotic relation between parent and offspring: the infant needs much assistance to survive, and the parent 'needs' to propagate genes. . . . A gene leading to cooing can interact with a gene leading to cradling to the benefit of both" (p. 17). Thus, if "the cat's purr can substitute for the coo, the cat has lucked into a good deal" (p. 17). So the cat may have evolved into a niche created by human need to nurture human infants.

I do not mean to suggest that such an evolutionary explanation can be extended to our responses to all other animals. For, many animals would be either potential predators or potential prey for humans. Failure to be able to distinguish them from human infants would confer a selective disadvantage. Still, what I think is most salient about Calvin's hypothesis is that it gives us a plausible example of a form of perception and emotional response that may be genetically programmed and is relatively crude and inaccessible to modification through learning or experience. Perhaps our general response to animals is also based on an innate but crude cognitive-emotional framework within which there is no clear differentiation between ourselves and other animals and in which we view animals as similar to ourselves (as having reasons, motives, thoughts, emotions). Hard centrists suspect as much and would view the process of educational maturation as a process of learning to disengage as much as possible from this framework, which is so uncalibrated in the very young.

Although it is plausible to speculate that our commonsense explanations of animal behavior are based on an innate framework of concepts for understanding others, both human and animal, I do not think this possibility provides a substantial case against anthropomorphic inferences.

First, our perceptual apparatus in general is innate (hard wired) and yet not for that reason to be discarded; we do not reject the validity of the reports of vision, for example, on these grounds. In addition, it is plausible to hold that our way of understanding each other is a system that is largely innate, rather than an inductive generalization built up from experience. [See Fodor and Chihara (1965) on the sort of theory we might have of other minds and for the argument that the theory can be partly innate, like a grammar, and is not based on criteria or symptoms defining M-predicates in terms of behavior.] So, unless we also wish to reject our ways of understanding each other, there is a problem in rejecting the system for understanding animals. The capacity itself is not necessarily defective or invalid in its operation. Second, as we grow up we *do* learn to make distinc-

tions between things that are really alive and things that are not, and between living things that really have mentality and those that do not. We do not for a moment think that dolls, raisins, or engines really have feelings. But common sense stubbornly refuses to draw a sharp line between humans and other animals and persists in retaining sympathetic feelings for animals and in understanding them along human lines in terms of M-predicates.

To demand that we draw a further line between humans and other animals such that we only apply our innate capacity to understand another mentality to fellow humans is to beg the question at issue. It cannot be that this capacity in itself produces invalid results, but rather that it is invalidly being “extended” to other creatures which it misconstrues. But to make that claim, a separate argument is required. It is not enough to speak of the evolutionary origins, the cognitive impenetrability of this capacity or the misconceptions it leads us to in childhood. To return to the cat, is it not *really* cuddly? Is it a mistake to find it cuddly, even if it is true that we have such a response because of the similarity of cats to infants? Obviously, to avoid begging the question the hard centrist must come up with a further argument to support his rejection of the application of this capacity to animals.

C. Anthropomorphic Inference Involves Invalid Methods

In addition to the hard centrist’s worry about the cognitive schemes with which we perceive and understand animals, philosophers have raised a related objection. Some of them have characterized the anthropomorphic inference in such a way as to make it dubious or even absurd. They have suggested that at bottom the inference requires a process of “projection.” Jonathan Bennett (1988) associates this view with Quine:

[Some philosophers] hold that we have no good *standards* governing our moves to conclusions about thoughts from premises about nonlinguistic behavior. . . . Quine, for instance, says that when we attribute beliefs to other animals we are speaking in a “dramatic idiom,” imaging ourselves in the animal’s shoes, so to speak, and saying on its behalf what we imagine *we* would think or be prone to say if *we* were barking at a cat up a tree or lunging at a toreador with our horns. (p. 200)

Bennett does not himself accept this view: “Despite recent sceptical literature tending the other way, I unashamedly conjecture that many animals have beliefs and wants, and that this is not anthropomorphism—not a ‘dramatic’ way of pretending to be a mouse or a gorilla” (p. 201). But as other influential philosophers obviously accept the idea that in ascribing thoughts to animals each of us is projecting ourselves onto the animal, we need to examine the tenability of this idea.

Although Bennett associates the idea of projection with anthropomorphism,

Quine (1960) does not use the term. He does, however, suggest that all ascriptions of propositional attitudes to fellow humans as well as to nonhuman animals involve projection, and he insinuates that projection is not in general an adequate method, and in the cases of animals, absurd:

Casting our real selves thus in unreal roles, we do not generally know how much reality to hold constant. Quandaries arise. But despite them we find ourselves attributing beliefs, wishes, and strivings even to creatures lacking the power of speech, such is our dramatic virtuosity. We project ourselves even into what from his behavior we imagine a mouse’s state of mind to have been, and dramatize it as a belief, wish, or striving, verbalized as seems relevant to us in the state thus feigned. (p. 219)

A Quinian hard centrist would no doubt base his or her case on the quandaries emphasized by Quine.

Like most of the ideas associated with the charge of anthropomorphism, projection has not been analyzed by those who casually invoke it to explain our thinking about animals. It therefore requires careful scrutiny. There are three points in particular that need investigation: (1) How is projection connected to anthropomorphism? (2) What is wrong with projection? Why does it lead to falsehood? (3) Does the anthropomorphic inference *depend upon* projection? Is it plausible to hold that projection is the mechanism that underlies anthropomorphic inferences? I will argue for two points. First, it is not clear that, if it is construed in a non-question begging way, projection necessarily delivers falsehoods. Second and more important, it is not at all obvious that anthropomorphic inferences depend upon projection. Projection is a red herring.

Imagining what I would do or think in the animal’s situation may be an unreliable method—but how unreliable and why? It is clear that if this is how I reason about what you will do or are thinking, I will get it wrong *some* of the time. But translated to the animal case, that would only yield situational anthropomorphism. Presumably what Quine and others have in mind is that it is always wrong or even entirely absurd. The hard centrist’s claim is that projection yields categorical anthropomorphism.

There is another ambiguity. In Quine’s (1960) view, projection underlies attribution of all propositional attitudes. For instance, “in indirect quotation we project ourselves into what, from his remarks and other indications we imagine the speaker’s state of mind to have been, and then we say what, in our language, is natural and relevant to for us in the state of mind thus feigned” (p. 218). But is projection supposed to underlie how I know in general that other people have M-states? Or is it only supposed to underlie how I arrive at beliefs about particular M-states, *given that I can correctly assume they have M-states*? If the latter, then translated into the animal case, this would again yield only situational anthropomorphism. To make a hard centrist case out of projection we must suppose that it is the former; that is, it is how we know at all that animals have M-states. To take this as an extension of the human case we have to accept the once-

common assumption in the philosophy of mind that on the basis of what I think and feel in particular circumstances, I ascribe to other humans thoughts and feelings in similar circumstances. I argue from my own case. If this is the role of projection, and projection is essentially defective when applied to animals, then any anthropomorphic inference that projection underlies becomes categorical anthropomorphism.

If projection works at all in the case of fellow humans, why can it not work for animals? If there is really such a mechanism of the imagination that enables each of us to overcome the required counterfactual (“if I were her”), why cannot the same mechanism overcome the counterfactuals involved in thinking about animals (“If I were Lassie”)? By hypothesis, I can abstract from sex, age, cultural, and physical differences in ascribing M-states to persons; why can I not go farther and abstract from greater physical differences? To deny that we can successfully project ourselves onto animals assumes that animals are too different from us for me to reason validly in this way. That assumption begs the very question at issue. Part of the absurdity of projection comes from the difficult counterfactual on which projection is supposed to be based. But if it is not absurd to reason this way in some cases, why does it become categorically impossible as soon as we reason about other animals?

Quine (1960) objects to projection because he views it as unscientific; he implies that the method of projection allows any answer and that it violates his strictures on evidence and explanation, which are purely behavioristic. [“In the strictest scientific spirit we can report all the behavior, verbal and otherwise, that may underlie our imputations of propositional attitudes . . .” (p. 219)] To Quine’s credit, this is why he is also suspicious of ascription of propositional attitudes to humans. But this is a standard that has been rightly displaced in scientific theory and practice. On behavioristic grounds commonsense explanations of other humans would be just as defective as are such explanations extended to animals.

The Quinian position has a further problem. In so far as we are to take seriously the idea that we only know from our own case, one can ask how Quine knows that projection, this mechanism of the imagination he claims structures our thinking about animals, is actually how we arrive at our ascription of M-predicates. Perhaps this is how Quine thinks about other people and perforce, other animals; but how does he know that anyone else does it this way? How could he have evidence for the truth of his own claim? Is this in fact an example of an invalid projection?

If in fact anthropomorphic inference does not depend on projection, then this whole case collapses. And indeed, I do not see any clear reason to suppose that this mechanism must underlie all anthropomorphic inferences and plenty of reason to deny it. For instance, much that is counted as paradigmatic of an-

thropomorphism cannot be plausibly understood as projection. If I find a baby elephant cute, how am I projecting myself?

No doubt we sometimes reason this way, at least about fellow humans, but when we do there is an important requirement. Consider a chess example: I may ask, what would I do in the situation of my opponent? Equally, I may reason: *I* would do X, but Karpov is the type who will do Y. Even to think, “suppose *I* were Karpov,” do I not have to know what Karpov is like, how he differs from me? In short do I not need a theory of Karpov, a model representing him? And in the case of animals (“suppose *I* were that elephant”) do I not need some theory of the animal, a modeling of what it is like to be the animal? But if so, projection loses its alleged function, for it only works if I already have some basis for understanding. Take the simplest example: next door a deer is munching on my neighbor’s grass. I immediately think: that deer must be hungry. I have not arrived at this thought by thinking: if *I* were to eat my neighbor’s grass, I would be hungry. If I went next door and started to munch on the lawn, I would perhaps have had a nervous breakdown or be suffering from a drug reaction. My reasoning was, rather, based on how I represent deer, as creatures who make grass a regular part of their diet.

We should contrast making ascriptions to the other based on understanding how the other responds mentally to a given situation with making them based on how *I* would respond in the other’s situation. What I think is absurd in the projection argument is the idea that I can only reason from my own case, that I know about myself and only about myself from the inside, and must somehow project myself onto the outside world to form any further inferences about people or animals. Such a picture has been convincingly shown to be a failure, however compelling it may seem if we accept the idea that each of us is an isolated Cartesian ego. On the one hand, the philosophical literature on the “other minds” problem demonstrates the thorough inability of the argument from analogy (the philosopher’s version of projection) to avoid total scepticism about other humans’ mental states. On the other hand, the private language argument, stemming from Wittgenstein (1953), has shown that we could not learn about our mental states from our own case. In particular we could not, I think, learn *what* our states are if we did not already know *that* other people have such states. It is not that one could never justifiably reason from analogy or by projection, but that this cannot be offered as the basis for how one knows that other humans have minds. Just as we do not make up our own language first and translate it into the public language, so we do not identify our own mental states first and then translate them into public discourse about mind. Instead, it is from learning the public language that we learn about M-states and that we have the M-states that others have. But if the projection mechanism is not the foundation for our *general* understanding of other humans, why insist that it is for our *general*

understanding of animals? (And to say less is to reduce the projection objection to saying that projection can lead to situational anthropomorphism.) Why suppose that projection must be the mechanism except for a commitment to the discredited view that “I know from my own case?”

A general dilemma reinforces this rejection of the claim that projection underlies the anthropomorphic inference. In so far as projection is understood as requiring an impossible thought (“if I were that deer”), and for that reason provides an invalid basis for anthropomorphic inferences, why suppose that those who ascribe M-predicates to animals are engaged in entertaining it? This convicts the person who ascribes thoughts to animals not just of irrationality, but of impossible thoughts. Moreover, it is not as if anyone has any direct evidence that other people who ascribe thought to animals are reasoning this way. If there is nonsense in the thought that constitutes projection, the nonsense is Quine’s, not that of those who do attribute M-predicates animals. They may be wrong, but they cannot be doing the impossible when they infer the animal’s M-state from its behavior.

The projection model of how we reason about others gains credibility only if there are no other conceivable possibilities. Yet other possibilities are conceivable. Rather than reasoning from our own cases, we could be born with an innate tendency to develop a theory of mentality that develops in parallel with our development of linguistic competence. We learn the theory primarily in application to people, but in addition we learn to apply it to higher animals on the basis of some of the same behavioral criteria that we apply in the human case. We learn such things as when we cry we are sad, when we strike out we are angry, when we romp around with each other we are playing, when we ignore food we are not hungry, and so forth. This theory may be mistaken if extended to nonhumans—and certainly it would be if we ignored the differences between us and them—but the question here is only whether we can think about animals without using the dubious mechanism of Quinian projection, and surely we can. If we “project” anything in thinking about animals it is human characteristics that we project upon them (rather than our individual qualities). This is projection in an entirely different and unproblematic sense. It may be incorrect, but it is far from impossible or even absurd. We have only to assume that animals are members of the class of creatures that have roughly similar responses to our own.

D. Anthropomorphic Inference Based on Inadequate Evidence

Finally I want to consider the claim that the charge of anthropomorphism is justified because of the inadequate evidence on which the anthropomorphic inference is based. Projection, if characterized as projecting M-predicates onto animals, may not be an incoherent notion, but there are those who would claim

that it is nonetheless inherently defective as a way of arriving at explanations of animal behavior. The particular concern I have in mind comes from the power of projection, or to use the terminology of John Bishop, “personification,” to apply to any behavior or form that resembles our own. This indiscriminate explanatory potential opens up the anthropomorphic inference to the criticism that it engages in an unwarranted inductive leap.

To make this into an argument that the anthropomorphic inference amounts to categorical anthropomorphism, that is, a categorical mistake, it must be argued that the evidential base of the inference is inherently defective. It fails to be adequate to the sort of inference it supports and it fails in a systematic way.

Such an argument is made by John Bishop (1980) in the course of discussing whether or not it is appropriate to give “rational” explanations of animal behavior; that is, explanations “in which phenomena are accounted for as the outcome of the agent’s practical reasoning from his beliefs and intentions” (p. 1). For Bishop the question of whether animals think is primarily the question of attributing practical reason to animals: Do animals act for reasons; do they behave as they do because they have goals and beliefs? He grants the plausibility of Hume’s inference: “aspects of dumb animal behavior bear striking phenomenal similarities to the kind of intelligent human behavior which counts as the very paradigm of intentional action—if such human behavior is explained as the outcome of practical reasoning, then the similar behaviour of languageless behavers may justifiably be similarly explained” (p. 4). Bishop maintains, however, that Hume’s inference—which is really just a subspecies of anthropomorphic inference—is merely a *prima facie* argument. We must beware, says Bishop, alluding to Davidson’s concern to avoid anthropomorphism, of what “is no more than habitual anthropomorphism about certain dumb animals” (p. 4). According to Bishop the Humean argument ignores

the possibility of what we may call the “personification challenge”: one can argue validly from the similarity between a dumb creature’s and a human’s behaviour to the conclusion that the former as well as the latter is the outcome of practical reasoning *only if* one can exclude the possibility that the similarity results from anthropomorphic projection. (p. 4, my emphasis)

If the personification challenge (PC) can be rejected, then Hume’s inference is satisfactory. Bishop contends, however, that the PC has yet to be overcome.

It is important to clarify what it means to describe Hume’s inference as only providing a *prima facie* case that rational explanations apply to animal behavior. Bishop wishes to argue not that Hume’s inference *is* mistaken, but that it *could* be mistaken. This leaves room for the success of arguments, like Davidson’s, that we cannot attribute beliefs to animals. Thus, although anthropomorphic inferences are compelling, they could still be wrong. If this is Bishop’s point, no conclusion I have drawn in this paper is inconsistent with it. Anthropomorphic inferences are at bottom inductive in that they apply a theory (the theory of M-

predicates) to explain animal behavior. And just as this theory *could* be incorrect as an explanation of human behavior, it could be incorrect as an explanation of animal behavior. But Bishop needs something stronger than this logical point. He needs to show a particular way the anthropomorphic inference could be mistaken; not mistaken in this or that case (situational anthropomorphism) but mistaken in general, yet not in such a way as to undermine our attribution of M-predicates to each other.

Bishop introduces the personification challenge by observing that there is a difference between behaving as if for a reason and *actually* acting for a reason. This is a distinction we make among human actions, but how does it apply to the present problem? One possible worry is whether we can tell when dumb animals have reasons and when they do not, and what their reasons are when they have them. These are the problems of avoiding situational anthropomorphism. The mistake involved in Hume's inference must rather be this: Possibly animals never have reasons, although they behave *as if* they have reasons. Why would we think this? What reason can be given for making such a distinction between human and nonhuman animals? We need an argument that establishes a category of creatures that never have any reasons (or beliefs and desires). It appears that Bishop thinks that our anthropomorphic inferences are based on a mental faculty that can project onto entities devoid of intentions the appearance of intentional action. But how would we know that these entities do *not* have intentions? Is there some fact of the matter that we have determined is misrepresented by "personification?" Surely we cannot say that we have direct access to the minds of dumb animals and we have found that they are empty.

Bishop's worry concerns our ability to make necessary distinctions on the basis of an animal's behavior: "But information-bearing belief-like states may play a role in the modification of a system's behaviour *without* its being true that the system acts for a reason. That kind of fiendish missile which modifies its track towards the moving target in accordance with information received through heat sensors provides an . . . apt example of such a system" (1980, p. 6). Davidson (1988), too, is concerned about such missiles:

someone might easily have no better or alternative way of explaining the movements of a heat-seeking missile than to suppose that the missile wanted to destroy an airplane and believed it could by moving in the way it was observed to move. This uninformed observer might be justified in attributing a desire and beliefs to the missile; but he would be wrong. (p. 477)

Does this just repeat the point that our inferences to M-predicates for animals are inductive in character, as are our inferences to M-predicates in humans, or is there a further problem concerning our capacity to make these inferences for animals? What is analogous to the missile in the case of animals? An argument that would do the required job goes like this: For all we know, animals are machines, and machines cannot act for a reason. Therefore, for all we know

animals cannot act for a reason. We know that animals are not machines, however, and anyway, why could a machine not act for a reason?

Maybe what is wrong with the fiendish missile is not that it is a machine but that it is such a simple machine. It modifies its movement, not because it has any recognizable states analogous to desires, needs, or beliefs, but because of a simple feedback loop. But in that case, we need not worry. Hardly any living organisms are this simple. Certainly a buzzard, to take an analogously behaving entity, does not merely and mechanically respond to the information it receives visually. [Richard Jeffrey (1985) doubts that even bees are analogous to Davidson's heat-seeking missile.] And on the neurological level at which we can analyze the causal mechanisms of a buzzard's vision, we can also analyze human vision as a confluence of complex causal processes.

What is wrong with the missile is that it is *nothing but* a system whose behavior is directly controlled by a simple modality of external information. A system that is thus controlled cannot be considered to have reasons for its behavior. This casts light on the nature of the PC. The fiendish missile is similar to a thermostat. Yet we would never attribute beliefs and desires to a house, whereas we might, momentarily, to a missile in flight, which looks like a bird. Hume's inference, then, may be acceptable as long as we make sure that we have not misperceived the entity in question, as we might a missile but would not a house.

Are we misled by the physical appearance of nonhuman animals in the way that we may be momentarily misled by the physical appearance of a missile? Surely not. Many animals have actual nervous systems, not something that imitates a nervous system. The buzzard really has an eye; the missile has something that looks like an eye. I am wrong if I think that the missile is a bird; I am not wrong if I think that the bird is a lot like me physically. More important is the perception of intentional action. I may mistakenly think the missile has intentions, desires, and so forth. This mistake could be put to rest by more observation of the missile's behavior, and therefore does not seriously challenge Hume's inference. Hume's inference gets off the ground only for creatures that appear on more than momentary inspection to behave like us.

What becomes of the personification challenge? There is a large class of entities, including humans, dumb animals, and machines, whose behavior can lead us to anthropomorphically infer that they act as a result of practical reason. Not all of these entities, however, do so. The challenge is this: Can we tell the difference? How do we know, for instance, that all animals are not just like the missile? The answer is that Hume's inference is designed to make just this distinction. We do have to exclude entities that cannot go wrong in their behavior, entities that cannot recognize, at least sometimes, that they are going wrong, and so forth. Why should we think that it is impossible to satisfy such criteria?

If it is a mistake to explain animal behavior as the result of beliefs and desires, the mistake must result from a failure of common sense to realize the

requirements of abstract philosophical theories concerning belief or action, for instance Davidson's requirement that to have a belief one must be a language interpreter. No argument for animal thought can be so decisive as to rule out the possibility that abstract philosophical theory may show beliefs or desires to require some condition that dumb animals do not satisfy. Our basic beliefs about human mental states are, in the same way, open to attack. Thus the possibility that animals do not have thoughts turns out to be similar to the possibility that commonsense views about human mentality, so-called "folk psychology," will prove incorrect (see Stich, 1983).

Do such possibilities show that our anthropomorphic thinking is guilty of the charge of categorical anthropomorphism? Not yet, for it has not been shown that our capacity to personify dogs can fall while our ability to personify each other can retain its validity. Still, for some thinkers there will be this difference: We can imagine that our theory (M-predicates) is entirely misapplied to animals, whereas, and contrary to the trendy rejection of folk psychology, we cannot imagine that theory being wrong in application to us. What I have argued is that no one has shown that we *can* imagine the former except by imagining the latter.

VIII. CONCLUSION

The concept of anthropomorphism is intimately tied to the negative criticism that it regularly implies. The idea that anthropomorphism names a widespread fallacy in commonsense thinking about animals is largely a myth, however, and the use of the term as a critical cudgel ought to be given up. It cannot stand for what it is supposed to. We have found that people commit no simple category mistake when they reason according to what I have termed the anthropomorphic inference. With the more complicated account of anthropomorphism developed above, we can say anthropomorphism exists, namely, situational and imaginative anthropomorphism. But we have found no adequate argument for the idea that commonsense thinking about animals instances the more fundamental mistake anthropocentrists commonly suppose it does: categorical anthropomorphism. Without a plausible defense that ascriptions of mental states to nonhuman animals is a categorical fallacy, the most basic assumption of critics of anthropomorphic thinking is shown to be untenable.

IX. REFERENCES

Asquith, P. J. (1986). Anthropomorphism and the Japanese and western traditions in primatology. In Else, J., and Lee, P. (eds.), *Primate Ontogeny, Cognition and Social Behaviour*, Cambridge University Press, Cambridge, pp. 61–71.

- Baier, A. (1985). Knowing our place in the animal world. Reprinted in her (ed.), *Postures of the Mind*, University of Minnesota Press, Minneapolis, pp. 139–156.
- Bennett, J. (1964). *Rationality*, Routledge & Kegan Paul, London.
- Bennett, J. (1988). Thoughtful brutes, Presidential Address, *Proceedings and Addresses of the American Philosophical Association*, Vol. 62, 197–210.
- Bishop, J. (1980). More thought on thought and talk, *Mind*, 89:1–16.
- Burghardt, G. M. (1990). Cognitive ethology and critical anthropomorphism: A snake with two heads and hognose snakes that play dead. In Ristau, C. A., and Marler, P. (eds.), *Cognitive Ethology: The Minds of Other Animals*, Lawrence Erlbaum, Hillsdale, New Jersey.
- Calvin, W. (1983). The lovable cat: Mimicry strikes again. In Calvin, W. (ed.), *The Throwing Madonna: Essays on the Brain*, McGraw-Hill, New York, pp. 14–21.
- Cassirer, E. (1950). *The Problem of Knowledge*, Yale University Press, New Haven.
- Carruthers, P. (1989). Brute experience. *J. Philos.* 86:258–69.
- Davidson, D. (1975). Thought and talk. In Guttenplan, S. (ed.), *Mind and Language*, Oxford University Press, Oxford, pp. 7–23.
- Davidson, D. (1988). Rational animals. In LePore, E., and McLaughlin, P. (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell, London, pp. 473–80.
- Edwards, P. (1967). Atheism. In Edwards, P. (ed.), *The Encyclopedia of Philosophy*, Macmillan, New York, pp. 174–189.
- Eltringham, S. K. (1982). *Elephants*, Blandford Press, Poole, Dorset.
- Fodor, J., and Chihara, C. (1965). Operationalism and ordinary language: A critique of Wittgenstein. *Am. Philos. Q.* 2:281–95.
- Hearne, V. (1987). *Adam's Task*, Random House, New York.
- Jeffrey, R. (1985). Animal interpretation. In LePore, E., and McLaughlin, P. (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell, London, pp. 481–87.
- Kendrick, W. (1988). Creepshow; John Crompton's wild kingdom. *Village Voice* June 28, p. 59.
- King, F. A., Yarbrough, C. J., Anderson, D. C., Gordon, T. P., and Gould, K. G. (1988). Primates, *Science* 240:1475–82.
- Midgley, M. (1978). *Beast and Man*, Cornell University Press, Ithaca, New York.
- Midgley, M. (1983). *Animals and Why They Matter*, University of Georgia Press, Athens, Georgia.
- Moss, C. (1988). *Elephant Memories: Thirteen Years in the Life of an Elephant Family*, Fawcett Columbine, New York.
- Newkirk, I. (1988). Published discussion in the Harper Forum, *Harpers Magazine* August, p. 43–52.
- Quine, W. V. O. (1960). *Word and Object*, MIT Press, Cambridge, Massachusetts.
- Ryle, G. (1949). *The Concept of Mind*, Barnes & Noble, New York.
- Shipley, J. T. (ed.) (1970). *Dictionary of World Literary Terms*, The Writer, Inc., Boston, Massachusetts.
- Stevenson, L. (1982). *The Metaphysics of Experience*, Oxford University Press, Oxford.
- Stich, S. (1983). *From Folk Psychology To Cognitive Science*, MIT Press, Cambridge, Massachusetts.
- Wittgenstein, L. (1953). *Philosophical Investigations*, Basil Blackwell, Oxford.